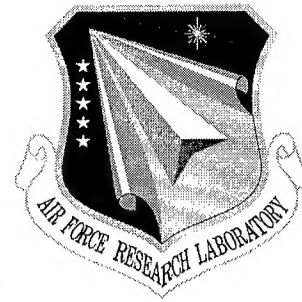AFRL-SN-RS-TR-1998-124
In-House Report
June 1998

HIGH CAPACITY FREE SPACE BACTERIORHODOPSIN BASE RECONFIGURABLE CROSSBAR

John E. Malowicki

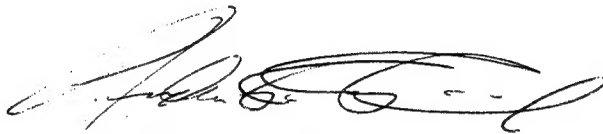*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

19980819 025

**AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-SN-RS-TR-1998-124 has been reviewed and is approved for publication.

APPROVED:

ANDREW R. PIRICH
Chief, Photonics Processing Branch
Sensors Directorate

FOR THE DIRECTOR:

ROBERT G. POLCE, Acting Chief
Rome Operations Office
Sensors Directorate

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
	June 1998	In-House	Oct 95 - Sep 97

4. TITLE AND SUBTITLE

HIGH CAPACITY FREE SPACE BACTERIORHODOPSIN BASED RECONFIGURABLE CROSSBAR

5. FUNDING NUMBERS

PE - 62702F
PR - 4600
TA - P1
WU - 24

6. AUTHOR(S)

Serey Thai

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/SNDP
25 Electronic Parkway
Rome NY 13441-4515

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-SN-RS-TR-1998-124

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/SNDP
25 Electronic Parkway
Rome NY 13441-4515

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-SN-RS-TR-1998-124

11. SUPPLEMENTARY NOTES

Air Force Research Laboratory Project Engineer: John E. Malowicki/SNDP/(315) 330-4682

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A novel bacteriorhodopsin based photonic crossbar system for broadband communications is proposed. This free-space dynamically reconfigurable NxN crossbar switch utilizes an intelligent holographic system for routing and switching by dynamically reconfigurable gratings of bacteriorhodopsin, which has high write/read photocyclicity that is greater than 10^6. The major advantages of the system include large interconnectivity density, transparent data redistribution, and fiber optic bandwidth capacity. Furthermore, the switching device resolves optical-to-electronic and electronic-to-optical conversion bottlenecks and reduces signal-to-noise degradation which is due to the conversions. This crossbar design is completely free of internal blocking which is one of the major drawbacks of guided optical crossbars. The system takes advantage of the parallelism and multidimensionality inherent in optics and can be scaled to a large capacity of NxN, while it offers the potential for low weight and portability which are a projected requirement for future broadband communications. In addition, a three laser beam technique is developed to investigate holographic diffraction characteristics of M-type phase holograms and transient gratings in a bacteriorhodopsin film. A peak diffraction efficiency is achieved by controlling the ratio of write to pump intensity. A maximized transient peak diffraction efficiency approaches a saturated value when the ratio is approximately unity.

14. SUBJECT TERMS

Bacteriorhodopsin Based Photonic Crossbar, Dynamically Reconfigurable Gratings, Transparent Data Redistribution, Fiber Optic Bandwidth Capacity, Transient Gratings

15. NUMBER OF PAGES

160

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ABSTRACT

A novel bacteriorhodopsin based photonic crossbar system for broadband communications is proposed. This free-space dynamically reconfigurable NxN crossbar switch utilizes an intelligent holographic system for routing and switching by dynamically reconfigurable gratings of bacteriorhodopsin, which has high write/read photocyclicity that is greater than 10^6. The major advantages of the system include large interconnectivity density, transparent data redistribution, and fiber optic bandwidth capacity. Furthermore, the switching device resolves optical-to-electronic and electronic-to-optical conversion bottlenecks and reduces signal-to-noise degradation which is due to the conversions. This crossbar design is completely free of internal blocking which is one of the major drawbacks of guided optical crossbars. The system takes advantage of the parallelism and multidimensionality inherent in optics and can be scaled to a large capacity of NxN, while it offers the potential for a low weight and portability which are a projected requirement for future broadband communications. In addition, a three laser beam technique is developed to investigate holographic diffraction characteristics of M-type phase holograms and transient gratings in a bacteriorhodopsin film. A peak diffraction efficiency is achieved by controlling the ratio of write to pump intensity. A maximized transient peak diffraction efficiency approaches a saturated value when the ratio is approximately unity.

ACKNOWLEDGMENTS

CONTENTS

List of Tables

List of Figures

Chapter 1

Current Switching Technology

Future communication networks will have throughput requirements exceeding the multi-Gbit/sec rate. The challenge for the designers of these networks will be development of nodes to perform the routing and switching functions of the network. Currently, the input-output interconnectivities in these nodes are performed either in the electronic domain or by guided optical crossbar switch.

Electronic nodes convert light from an incoming fiber into an electronic signal. After routing and switching functions are performed in the electronic domain, the electronic signal is reconverted to light for transmission over a fiber. The inefficiencies of the optical/electrical and electrical/optical converters combined with the low transmission bandwidth of the electronic switches limit the performance of the network; reducing bandwidth, attenuating the signal and adding noise. The bottlenecks created by the optical/electrical and electrical/optical converters have been massively reduced by guided optical crossbar. It is a space division switching network that routes M-incoming links to N-outgoing links. The crossbar, however, is not a blockfree device. It does not resolve data traffic congestion of either the internal or external blocking type. It does, however, route data in the optical domain which in itself is a major improvement over its electronic counterpart.

This chapter introduces the fundamental principles of the guided optical crossbar. Sections 1.1 and 1.2 provides a brief introduction to the crossbar switch and its history. Section 1.3 is devoted to optical fiber transmission. Section 1.4 covers the basic properties of a 2x2 directional coupler switch. Guided photonic crossbar switches are described in Sec. 1.5. The motivation for a free space optical crossbar is outlined in Sec. 1.6, and the concept of a free space crossbar switch is introduced in Sec. 1.7.

1

1.1 An Introduction to Switch Technology

We are living in a society where exchanging information is a vital part of our daily routine. Exchanging information between two agents is unanimously called communication. A communication system is used to convey messages over a distance beyond that which naked eyes and ears can detect. A basic communication system consists of a transmitter, receiver, and guided or unguided transmission medium. An unguided medium is a means to transmit electromagnetic waves in a free space such as air, vacuum and/or seawater. A guided medium is a physical path in which electromagnetic waves propagate such as twisted pair, coaxial cable, waveguide and optical fiber.[1] Optical fiber is probably one of the greatest technical developments in the 20th century. This medium is now commonly used in optical communications. Optical communications has been with us since the dawn of time. For instance, mankind has used hand, smoke or fire signals to convey messages from senders to receivers. The information being transmitted is carried by the sun's radiation which is detected by eyes and the image is processed by the brain into a message.[2] Sometimes the transmitted messages need to be subjected to a series of repeaters before it reaches a destination at a certain distance away from a source. As time has progressed, communication systems have become more sophisticated in order to keep up with user's demands. In the 1790s, a French engineer named Claude Chappe invented the optical telegraph.[3,4] His system was used to relay messages to a destination through a series of human operator semaphores mounted on towers. Almost a century later, in 1880, Alexander Graham Bell patented his invention for a photophone which was an optical telephone system.[5] The system required sunlight as a transmission medium. The message was transmitted by a thin, voice-modulated mirror to a photoconducting selenium cell receiver which in turn converted the message to electrical current. However, his main contribution is credited to his most practical invention : the invention of the telephone which was patented in 1876.[6] His invention gave birth to the telephone industry and created a very crucial issue right from the onset - the issue of interconnection. Establishing a direct

link from any phone to any other phone requires n(n-1)/2 channels for n phones since each phone has n-1 incident channels. It is more practical to converge all wires on a switching center where individual phone-to-phone connections would be made.[7] The interconnection of n phones would then be minimized to n channels. Therefore, a switching system is a vital component in a communication system.

The first commercial telephone exchange was established in Boston in May 1877 as an adjunct to a burglar alarm system, and this was followed by the first commercial telephone service as an independent entity which opened in New Haven, Connecticut in January 1878.[6] It was not until about seventy-five years later that mainframe computers were first introduced to commercial applications. The development of computers may have evolved from the abacus but their applications are endless. After mainframes were commercially introduced in 1953,[8] computers have become smaller and smaller and their applications have spread across the spectrum from personal to business. Hence, the information technological age has induced this hardware (telephone and computer) into a common household product in the United States. Since audio and video data are digitizable, they can be standardized to the same binary format as computer data to transmit on the same line, or more importantly, to be routed by the same switch. The merger of telephone and computer communication has created an enormous burden on the switching network due to heavy traffic congestion. This situation can be rectified by utilizing intelligent crossbar switches and increasing the carrier frequency to an optical domain which will produce a larger transmission bandwidth and, in turn, a larger information capacity.

1.2 A Brief History of Crossbar Switch

Crossbar switches are deeply rooted in a wonderfully rich history of telecommunication. The earliest switching systems were manually attendant switching

3

systems where an operator at a switchboard interconnected a requested line for a customer. An alerting lamp at a switchboard was illuminated after a handset was lifted from a telephone to cue switchboard operators, and the operator manually plugged a cord into jacks to establish a link according to the request of a customer. Switching innovations next evolved to an automated system invented by Strowger.[7] This system was the automatic use of step-by-step mechanical switches. The switches move in two dimensional steps.[9] If a handset is lifted, the line finder switch moves first vertically and then horizontally in a two dimensional array to locate the line so that the first selector can send a dial tone back to the handset. Thereafter, the first selector advances step-by-step vertically to the level corresponding to the first number dialed and then horizontally to determine an available connection. The same process repeats itself for the following dialed digit with another selector. A busy signal will be sent back to the handset if any selector was unable to determine a connection availability. The last two digits of a phone number signal the connector to step first vertically and then horizontally to the contacts to complete the connection, and delivers a ringing signal to the called phone through the connector.

The wheel of switch evolution soon necessitated switch designers to confront the issues of quick switching responses and larger accessibility. The designers came up with the Panel System which was an improvement on the Step-by-Step System. This system increased accessibility of switching elements and introduced routing control between the line-finder and the first selector.[9] This new system can be viewed as an first attempt to separate the control function from the interconnecting function. This switching philosophy was successfully incorporated into a next generation of switch design. The mechanical crossbar system was developed in the late 1930's and the early 1940's.[9] The basic operation of the system is as follows: the interconnection was done by operating a horizontal and vertical magnet. It was the first system to implement a complete and successful separation between the switching matrix interconnecting function and routing

control function. This basic philosophy is still utilized in the main stream of today's switching design for long haul communications.

A growing complexity of mechanical switches has challenged switch designers to explore alternative possibilities. The Director of Research at Bell Laboratories, Mervin Kelly, had an intuitive feeling as early as 1936 that one day the mechanical means of providing interconnections would have to be replaced by electronic switches.[10] The actual development of the electronic crossbar was not realizable until after the invention of the transistor at Bell Laboratories in Murray Hill, New Jersey on 23 December 1947.[10] This major invention of the transistor was credited to three distinguished researchers - John Bardeen, Walter Brattain and William Shockley who were later (1956) awarded the Nobel Prize in Physics for their efforts.[10]

The invention of the transistor did not only make it possible to develop stored program control of switching systems for routing information but also made it possible to design electronic digital crossbar switching systems which could interconnect signals in digital form. The crossbar network is a nonblocking architecture with a very straightforward routing algorithm. The development of the digital crossbar allows computers to network one another via telephone lines. Data transmission over the same line for both telephone and computer application causes data congestion at each node of the network. This traffic congestion has motivated switch designers to explore a new realm of possibilities. The invention of the laser and optical fiber made it possible to route this data in the optical domain. Utilizing optical switches will eliminate bottleneck problems caused by the conversion of transmitted information from the optical domain to the electrical domain for switching operation, and vice versa to retransmit over a fiberoptic cable. In addition, this switching system will allow the users to fully capitalize on all the benefits of optical data transmissions. However, designing a practical optical switching network is one of the major challenges that has been pursued for quite some time by the switching design community.

1.3 Optical Fiber Transmission

Ever since ancient times, various forms of communication systems have developed over the years. The driving principal motivations were always the same: to enhance transmission bandwidth, increase data rate, extend transmission distance with very low attenuation, minimize error rate and optimize transmission fidelity. Although various spectral regions of the electromagnetic spectrum are utilized as an information carrier from source to sink, optical fiber at infrared carrier frequencies provides very low insertion loss, very low cross talk and high bandwidth. In addition, fiber is very lightweight, immune to electromagnetic interference and relatively intruder resistant. These advantages of optical fiber transmission have led to world-wide installation of commercial fiber cables for all communication purposes beginning in the early 1980's. A basic fiber transmission link comprises a laser as a transmitter, fiber optic cable as a transmission medium, and photodetector as a receiver as shown in Fig. 1-1.

The laser emits a coherent light source as an optical carrier, and signal modulation is either done by direct internal modulation or external modulation. An example of external modulation is a spatial light modulator which modulates a CW laser beam by superimposing the information signal onto the electromagnetic wave. Light is an electromagnetic carrier wave which is governed by the same theoretical principles as all forms of electromagnetic radiation.

Fiber is a very high fidelity guided transmission medium that bridges a transmitter to a receiver. A photonic cross-bar network simplifies the complexity of the cabling system and still maintains the highest degree of effectiveness to route a signal to any desired destination with the minimum of cable lines, not to mention the economic sense of switch application.

At the receiver end, a photodetector converts optical radiation into an output electrical current or voltage with the original signal integrity remaining intact. Then a

receiver performs the necessary tasks to demodulate the signal including amplification, decoding, and/or other signal processing.

ELECTRICAL INPUT SIGNAL   Photonic cross bar   ELECTRICAL OUTPUT SIGNAL

LASER — SLM — Fiberoptic cables ⊠ Fiberoptic cables — PHOTODETECTORS
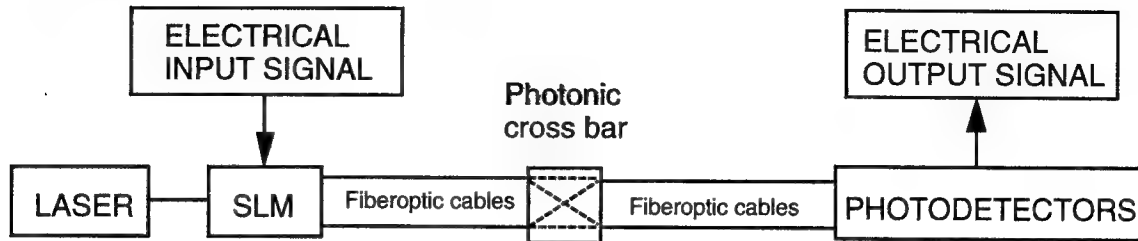
Fig. 1-1 A basic optical fiber transmission link.

The maser was invented by Townes in 1954, and it is an acronym for Microwave Amplification by Stimulated Emission of Radiation.[11] Four years later (1958), Schawlow and Townes proposed the possibility of applying the maser concept to the optical frequency range. But it was not until 1960 when Maiman historically demonstrated the successful operation of the first pulsed ruby laser.[12] This laser emitted, for the first time, a coherent brilliant red light that was not found in nature. Lasers and masers are both governed by the same population inversion process. The laser is a device that generates or amplifies coherent radiation at optical frequencies. Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. It was initially known as optical maser because it operates on the same general principle originally invented at microwave frequencies. The operation frequencies of lasers range from infrared to visible including the ultraviolet region of the electromagnetic spectrum.[13] Laser technology, as in other technologies, follows an evolutionary process. The successful demonstration of the ruby laser soon spun off gas and semiconductor lasers. Early in 1961 the successful experimental results of the first continuous wave gas laser using helium and neon was announced by Javan et al.[14] GaAs semiconductor lasers were successfully developed in 1962 by three independent groups consisting of Nathan et al.,[15] Hall et al.,[16] and Quist et al.[17] Thereafter, many different

materials with laser oscillation potential have been discovered, and several historical breakthroughs have been made to improve laser technology. The early laser research conducted intensively in the United States resulted in the first successful operation of the solid-state ruby laser. The discovery of the laser gave birth to a next generation of free space optical communication. Instead of using signal by fire, reflecting mirrors, or signaling incandescent lamps, the next generation light source became a monochromatic wave with phase-coherent polarization that was not found in nature. This monochromatic light source introduced an endless potential for system applications. However, atmospheric transmission presents a great limitation to long haul communications. It is restricted to a line of sight, and it is vital to operate at a transmission wavelength in the atmospheric window which gives low absorption. Furthermore, the reliability of the laser link is severely affected by the disturbances of elements such as dust, fog, rain, snow and atmospheric turbulence. For example, an optical beam under the influence of rainy or foggy weather can suffer great absorption and scattering - as much as 20 to 30 dB/km in worst conditions.[18] This obstacle has led researchers to explore new realms of possibilities. The efforts have resulted in a major breakthrough with the development of almost lossless optical fiber transmission medium.

The history of fiber may have begun with a need to explore inside the human body.[3,4] A medical student in Munich named Heinrich Lamm was perhaps the first person known to report on an image transmission of a light bulb filament through a short and unclad fiber bundle in 1930. Abraham van Heel of the Technical University of Delft in Holland reported in 1954 on a short distance transmission of light through bare fiber bundles. Subsequently, he was enlightened by an American optical physicist named Brian O'Brien that light inside fiber could be guided by cladding. Van Heel was the first person to experiment on a fiber with a lower refractive index transparent cladding. The glass-clad fibers were first developed by Lawrence Curtiss working with physician Basil Hirschowitz and physicist Wilbur Peters at the University of Michigan to make a fiber optic

endoscope for internal stomach examination around 1957. In early 1960, the attenuation of the glass-clad fibers was about 1 dB/m which was still much too high for a communications transmission medium. Detailed theoretical and experimental studies on a circular cross-section optical fiber were reported by Dr. K. C. Kao et al. in July 1966.[19] The studies focused on loss characteristics and information capacity for communication purposes. The investigation of information capacity concentrated on mode stability, dispersion, and power handling of an optical fiber. According to Dr. Kao, fiber is relatively strong and bending loss is negligible until it reaches a certain critical radius of curvature. He also observed that fiber loss was mainly due to impurities in glass. He stated that "it is foreseeable that glass with a bulk loss of about 20 dB/km at around 0.6 μm will be obtained, as the iron-impurity concentration may be reduced to 1 part per million." In addition, he concluded that "the refractive index of the core needs to be about 1% higher than that of the cladding. This form of waveguide operates in a single HE_{11}, E_0 or H_0 mode and has an information capacity in excess of one gigacycle per second." Dr. Kao's forecast was soon realized by a group of researchers at the Corning Glass Works (now Corning, Inc.). They were Robert Maurer, Donald Keck and Peter Schultz. In 1970, they successfully developed low loss, single mode fibers by fused silica, and added dopants to make the core index slightly higher than that of the cladding.[3,4] Corning optical fiber had a core diameter of 3 to 4 μm and the cladding diameter was approximately 60 times larger than that of its core. Corning reported optical fibers having attenuation below 20 dB/km measured at the 632.8 nm helium neon line wavelength.[20] Corning's breakthrough opened the door to fabricate nearly lossless fiberoptic cables. The attenuation of optical fiber is optical wavelength dependent. Fiber loss at 850 nm is about 2 dB/km, and is as low as 0.4 dB/km at 1.3 μm wavelength. The fiber attenuation reaches 0.2 dB/km at a wavelength of 1.55 nm.

A photodetector is a device that converts incoming optical radiation to output electrical signals. The detector is a square-law device. Thus, the photocurrent is

proportional to the incident power, or equivalent to an average of the square value of the electric field over one optical period. In addition, the properties such as responsivity, spectral response, and rise time are important characteristics of a detector. Thermal and photoelectric detectors are the two commonly used classes of photodetectors.[2,29] Converting photon energy into heat is the basic operation principle of a thermal detector, but it has relatively slow response and is inefficient. On the other hand, photoelectric detectors are most suitable devices for optical communication purposes. Its principle of operation is based on the external and internal photoelectric effect. The external photoelectric effect is found in photomultipliers, vacuum photodetectors, and phototubes where incident photon energy excites electrons over the potential barrier of a material surface into the vacuum and they become free electrons.[21] The internal photoelectric effect is found in many modern semiconductor photodetectors where absorbed photons generate electron-hole photogeneration and they remain within the material. A photon raises an electron from the valence band to the conduction band creating a free mobile charge carrier by breaking the semiconductor-semiconductor bond with an energy greater than the energy of the band gap; the band gap, or forbidden gap, is a gap between the valence and conduction band.[22] The most important process in internal photoelectric effect of a semiconductor photodetector is photoconductivity.

History has revealed that Mr. May and Willoughby Smith were the first persons to discover a photoconductive effect in selenium and reported the results to the Society of Telegraph Engineers on February 17, 1873. Alexander Graham Bell was the first person to utilize selenium as a photoconducting receiver in his invention of the photophone.[5,23] It is remarkably interesting to note that photoconductive semiconductor detectors are the backbone optical receivers in the present information superhighway technological age, so called "Cyberspace".

1.4 A 2x2 Directional Coupler Switch

A conventional directional coupler consists of two parallel waveguides where their active interaction regions are in close proximity. This mechanism can be used to transfer the light energy from one waveguide to the other via the affect of evanescent field coupling, and this transfer can be controlled electrically.[24] One application is to employ this device as a 2x2 guided optical crossbar switch.

The monochromatic plane waves travel in an arbitrary positive direction, according to the engineering convention, and can be represented as follows [25,26] :

$$E(r,t) = E_0 e^{-i(k \cdot r - \omega t)} n \tag{1-1}$$

where n is the direction of the polarization vector. The time of wave propagation is t, and ω is the angular frequency of the propagated wave. This frequency is $\omega = 2\pi f$, where f is the propagation frequency. A radius vector r of a propagated wave is $r = x a_x + y a_y + z a_z$ in the rectangular (Cartesian) coordinate system while a_x, a_y, a_z are the unit vectors in the direction of propagation of the three coordinate axes x, y, and z, respectively. A wavenumber vector k is defined as $k = k_x a_x + k_y a_y + k_z a_z$. The complex phasor quantities of the electric fields is $E_o = |E_o| e^{i\delta}$, where $|E_o|$ is the amplitude of the wave. It is a positive number, representing the maximum displacement from equilibrium. The phasor angle δ is called the phase constant.

Consider two optical waveguides lying parallel in the z-axis as shown in figure 1-2. The electric field of Equ. (1-1) can then be modified to represent the fields in both waveguides as follows:

$$E_1(z,t) = E_{o1} e^{-i(k_1 z - \omega_1 t)} n \tag{1-2}$$

11

$$E_2(z,t) = E_{02}e^{-i(k_2z-\omega_2t)}\mathbf{n} \qquad (1\text{-}3)$$

Let's consider the coupling effect between both arms of a coupler to be insignificant because the effect is too weak. Then the propagation variation over a changing distance along both waveguides 1 and 2 is simply the uncoupling equations of modes 1 and 2, respectively.[27] The functional form expressions are :
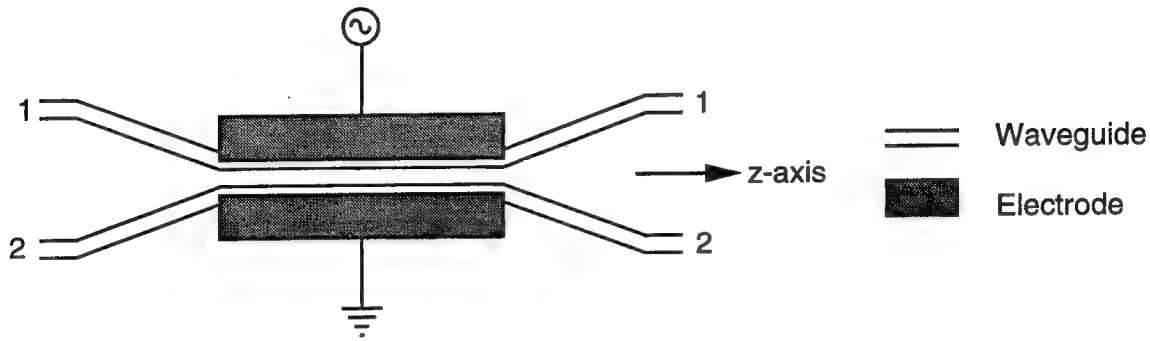
Fig. 1-2 A 2x2 directional coupler.

$$\frac{dE_1}{dz} = -ik_1E_1 \qquad (1\text{-}4)$$

$$\frac{dE_2}{dz} = -ik_2E_2 \qquad (1\text{-}5)$$

If the two waves of modes 1 and 2 are weakly coupled such that the optical energy in one waveguide is effected by the energy from the adjacent waveguide, and vice versa, the expressions representing two coupling modes become :

$$\frac{dE_1}{dz} = -ik_1E_1 + M_{12}E_2 \qquad (1\text{-}6)$$

$$\frac{dE_2}{dz} = -ik_2E_2 + M_{21}E_1 \qquad (1\text{-}7)\ .$$

where M_{12} and M_{21} are the coupling coefficients from waveguide 1 to waveguide 2, and from waveguide 2 to waveguide 1, respectively.

The coupling principle of an optical waveguide directional coupler is analogous to the principle of a mutual inductance circuit, where a current flowing in one coil creates a magnetic flux not only about that coil, but also about a second coil that is in close proximity. Hence, it produces a voltage across the terminals of the second coil by its time-varying flux surrounding.

The coupling coefficients can be determined by invoking the principle of total energy conservation. The sum of the optical power between two adjacent codirectional waveguides must be equal to a constant. Accordingly, $\left|E_1\right|^2 + \left|E_2\right|^2 = C$, where C is a constant. Therefore, the rate of changing total power with respect to a changing distance must go to zero. Thus, the mathematical representation is :

$$\frac{d}{dz}\left(\left|E_1\right|^2 + \left|E_2\right|^2\right) = 0 \tag{1-8}$$

Each term in Equ. (1-8) clearly contains a sum of two derivative conjugated pairs which can be written in detail as follows:

$$\frac{d}{dz}\left(\left|E_1\right|^2\right) = E_1 \frac{dE_1^*}{dz} + E_1^* \frac{dE_1}{dz} \tag{1-9}$$

$$\frac{d}{dz}\left(\left|E_2\right|^2\right) = E_2 \frac{dE_2^*}{dz} + E_2^* \frac{dE_2}{dz} \tag{1-10}$$

where the asterisk "*" represents a complex conjugate. The right hand side of the above equations can be further written as a combination of a nonderivative multiplication sum by substituting Equ.(1-6) and Equ.(1-7) into Equ. (1-9) and Equ. (1-10). Notice that both equations (1-9) and (1-10) contain a first order derivative conjugated traveling wave. Hence, the conjugation operation must be performed on both equations (1-6) and (1-7) for

substitution. Applying the resultant solutions to Equ. (1-8) expands the left hand side of this equation to a zero-order derivative combination which can be written as follows:

$$\frac{d}{dz}\left(\left|E_1\right|^2 + \left|E_2\right|^2\right) = E_1^* E_2\left(M_{12} + M_{21}^*\right) + E_1 E_2^*\left(M_{12}^* + M_{21}\right) = 0 \quad \text{(1-11)}$$

Note that each term in the preceding equation is a complex conjugate of each other. Since a product of two traveling waves between modes 1 and 2 cannot vanish, then the sum of the coupling coefficients must be zero for Equ. (1-11) to be valid. This yields a condition causing mutual coupling coefficients to become $M_{12} = -M_{21}^*$ and $M_{21} = -M_{12}^*$. Therefore, the principle of the total energy conservation imposes a restriction on coupling coefficients M_{12} and M_{21}.

The solutions of Equ. (1-6) and Equ. (1-7) provides considerable insight into the behavior of the monochromatic plane wave as it propagates through a waveguide coupler. One can approach this problem by treating Equ. (1-6) and Equ. (1-7) as first-order linear differential equations. Let's assume that $E_1(z,t)$ and $E_2(z,t)$ are proportional to each other. Thus, both equations become homogeneous differential equations of degree 1. Hence, the solutions are in the form $E_1 \propto e^{-ikz}$ and $E_2 \propto e^{-ikz}$. Substituting both exponential solutions into Equ. (1-6) and Equ. (1-7) leads to the following equations :

$$i\left(k - k_1\right)E_1 + M_{12}E_2 = 0 \quad \text{(1-12)}$$

$$i\left(k - k_2\right)E_2 + M_{21}E_1 = 0 \quad \text{(1-13)}$$

Both fields E_1 and E_2 can be eliminated by solving both equations simultaneously. The results can be expressed as a quadratic form :

$$k^2 - k\left(k_1 + k_2\right) + k_1 k_2 - \left|M_{12}\right|^2 = 0 \quad \text{(1-14)}$$

This quadratic equation yields two roots, one at k_a and other at k_b as follows:

14

$$k_{a,b} = \frac{k_1 + k_2}{2} \pm \sqrt{\left(\frac{k_1 - k_2}{2}\right)^2 + |M_{12}|^2}$$ (15)

At this juncture, let 's express the most general solutions to Equ. (1-6) and Equ.(1-7) in terms of four arbitrary constant coefficients:

$$E_1(z) = Ae^{-ik_a z} + Be^{-ik_b z}$$ (1-16)

$$E_2(z) = Ce^{-ik_a z} + De^{-ik_b z}$$ (1-17)

Notice that time t has been omitted from all previous equations for the sake of simplicity because it is an independent constant with respect to variable z. The coefficients A, B, C, and D are arbitrary constants which can be determined if appropriate boundary condition values are known. Let's consider the initial boundary values at z=0 to be $E_1(0)$ and $E_2(0)$. Utilizing these boundary conditions makes it possible to write four new equations, two of which are directly from Equ. (1-16) and Equ. (1-17) with z=0. The other two equations are found by substituting Equ. (1-16) and Equ.(1-17) into Equ. (1-6) and Equ. (1-7), and invoking the boundary conditions at z=0. The four new equations are :

$$E_1(0) = A + B$$ (1-18)

$$E_2(0) = C + D$$ (1-19)

$$-ik_a A - ik_b B = -ik_1(A + B) + M_{12}(C + D)$$ (1-20)

$$-ik_a C - ik_b D = -ik_2(C + D) + M_{21}(A + B)$$ (1-21)

These new equations are the vital ingredient to a successful determination of constant coefficients A, B, C and D in terms of known variables. After substituting Equ. (1-18) and Equ. (1-19) into Equ. (1-20) and Equ. (1-21), along with some algebraic manipulations, these coefficients are evaluated in closed form as follows :

$$A = -\frac{1}{2\Omega}\left[(k_b - k_1)E_1(0) - iM_{12}E_2(0)\right]$$ (1-22)

15

$$B = \frac{1}{2\Omega}\left[(k_a - k_1)E_1(0) - iM_{12}E_2(0)\right] \qquad (1\text{-}23)$$

$$C = -\frac{1}{2\Omega}\left[(k_b - k_2)E_2(0) - iM_{21}E_1(0)\right] \qquad (1\text{-}24)$$

$$D = \frac{1}{2\Omega}\left[(k_a - k_2)E_2(0) - iM_{21}E_1(0)\right] \qquad (1\text{-}25)$$

where $\Omega = \sqrt{\left(\dfrac{k_1 - k_2}{2}\right)^2 + \left|M_{12}\right|^2}$. It is interesting to note that if the ratio of n_1/λ_1 is

equivalent to the ratio of n_2/λ_2, then the real propagation constants $k_1 = k_2$, and Ω is proportional to the coupling coefficients M_{12}. This condition implies that the difference between two coupled modes in the waveguide is 2Ω or $2M_{12}$. The refractive index and propagation wavelength are represented by the conventional symbols n and λ, respectively.

Replacing the constant coefficients of equations (1-22) through (1-25) into Equ. (1-16) and Equ. (1-17), and invoking Equ. (1-15) to simplify the expressions, the resultant solution is as follows :

$$E_1(z,t) = \left[(\cos\Omega z + i\frac{\Delta\beta}{2\Omega}\sin\Omega z)E_1(0) + (\frac{M_{12}}{\Omega}\sin\Omega z)E_2(0)\right]e^{-i(\beta z - \omega t)} \qquad (1\text{-}26)$$

$$E_2(z,t) = \left[(\frac{M_{21}}{\Omega}\sin\Omega z)E_1(0) + (\cos\Omega z - i\frac{\Delta\beta}{2\Omega}\sin\Omega z)E_2(0)\right]e^{-i(\beta z - \omega t)} \qquad (1\text{-}27)$$

where $\beta = (k_1 + k_2)/2$ is a propagation constant, and $\Delta\beta = k_2 - k_1$ is a mismatched propagation constant between the two modes.

Graphic illustration of mode intensities $|E_1(z,t)|^2$ and $|E_2(z,t)|^2$ will elucidate how a codirectional coupler can be used in switching applications. Let's consider a perfect phase matching condition where $\Delta\beta = 0$, and waveguide 1 is the only mode that is fully excited at

16

z = 0 ($E_1(0) = 1$ for a normalized condition). Thus, there is no wave propagation in waveguide 2, initially ($E_2(0) = 0$). These assumptions lead to the computer simulation as shown in Fig. 1-3. It demonstrates how a coupler can be used to switch optical power from a straight-through state where light traveling appears at the output of the same waveguide as it has entered, to a complete crossover state which shifts all power from waveguide 1 to waveguide 2.[28] Remember that wavenumber k_1 must be equal to wavenumber k_2 in order to achieve a complete crossover condition. The complete crossover state occurs at $\Omega z = \pi/2$, and Ω is simply a coupling coefficient M_{12} in this case.

Hence, a distance $z = (\pi/2)M_{12}$ is called the transfer distance or coupling length.[29] These wavenumbers can be controlled by the electro-optic effect which is induced by an applied electric field. A directional coupler is two waveguides lying parallel a small distance, d, apart on an electro-optic substrate such as $LiNbO_3$ and sandwiched by electrodes. Applying dc voltage V to the electrode induces a positive dc electric field E on one waveguide and a negative electric field -E on the other waveguide since electric field is equivalent to the applied voltage divided by the distance d ($E \approx V/d$). The induced electric field E alters the refractive index causing it to increase in one waveguide and decrease in the other. The refractive index changes influence the phase mismatch of the propagating waves in the coupler. It can also be conversely stated that the phase mismatch is simply proportional to the difference between the refractive indices Δn in a directional coupler. Its functional expression is $\Delta \beta = (2\pi/\lambda_o)\Delta n$, where the changing refractive index Δn is simply an index difference between both modes ($\Delta n = n_2 - n_1$). The net refractive index difference Δn can be determined by solving an equation of the index ellipsoid.[30] The solution is generally in the form $\Delta n = 1/2n^3 r_{ij} E$, where n is the principal refractive index, r_{ij} is the linear (Pockels) electro-optic coefficients, and the indices i and j are the contracted notation

17

of the electro-optic tensor. Therefore, the phase mismatch $\Delta\beta(V) = k_2 - k_1$ can be controlled by applying voltage. In other words, a directional coupler switch is governed by two important parameters - the coupling coefficient M_{12} and the phase mismatch of the wavenumbers $\Delta\beta(V)$.
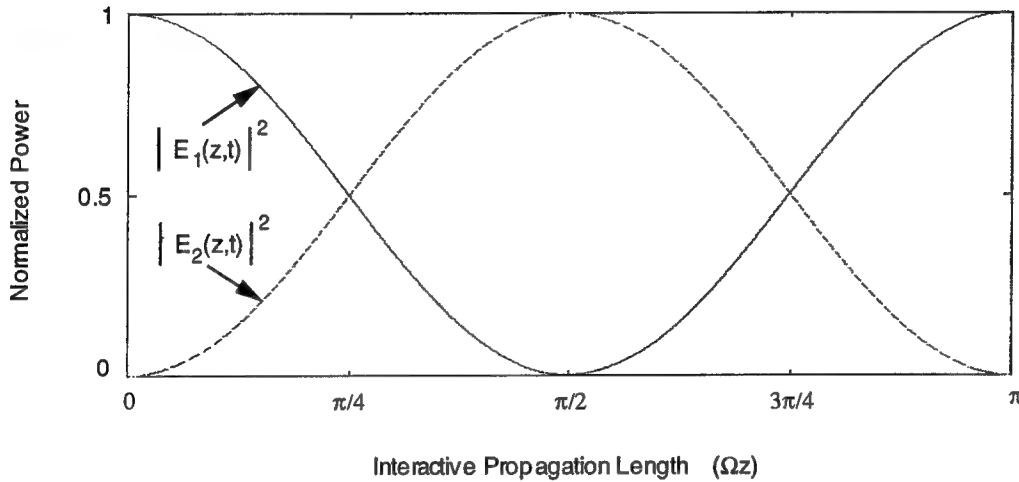
Fig. 1-3 Phase-matched power transfer of a codirectional coupler.

1.5 Guided Photonic Crossbar

The guided crossbar architecture is a space division switching network that routes M-incoming links to N-outgoing links. This network is not a blockfree system. It does not resolve data traffic congestion from either internal or external blocking. It does, however, route data in the optical domain which in itself is a major improvement over its electronic counterpart. The crosspoints of the crossbar are MN which are the product of input ports to output ports. A crosspoint is simply a 2x2 switch that processes a straight or cross state. Hence, a large-array crossbar configuration is normally constructed based

18

on 2x2 cascaded switches. If an input to an output link has NxN dimensions, then the complexity growth rate of a crossbar switch is on the order of N^2 ($O(N^2)$) crosspoints.

Optical fibers offer virtually unlimited transmission bandwidth which can be used to accommodate the increasing demand for high-data-rate communication systems. On the other hand, electronic switches are a very mature technology but they can not match fiber optic transport capacities. Therefore, switch designers have concentrated their research and development efforts on photonic switches to alleviate the switching bottleneck conversion problem for a broadband communication network. Making a transition from electronic crossbar switches to photonic crossbar switches is easily realizable because a basic crossbar network topology is applicable to either domain, and a maturing directional coupler technology (commercially available) makes it possible to employ this coupler as an optical crosspoint. Using a directional coupler as a 2x2 switch will serve two operational

straight-through state   crossover state
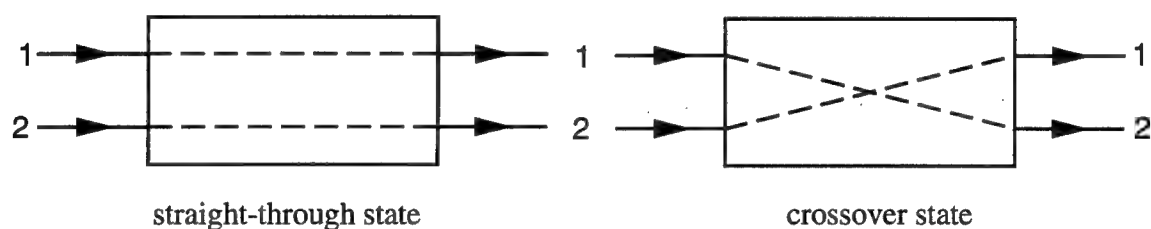
Fig. 1-4 The functional states for a 2x2 crossbar configuration.

functions, namely, a "crossover" and a "straight-through" states. The symbolic representation of these two states is put forth in Fig. 1-4 where a control function is not graphically illustrated for simplicity sake, but the reader should be well aware of its existence. A coupler can be driven by a dc field to a straight-through state to interconnect input channel 1 and 2 to output channel 1 and 2, respectively. It can also be toggled to a crossover state to reconfigure interconnection from input channel 1 to output channel 2, and vice versa at one switching cycle. If data flow conforms to these two conditional states

19

then a coupler is said to be in a nonblocking mode. On the other hand, 50% of the time both input ports of a coupler could request access to the same output port at the same time. This operational request presents an internal blocking conflict. Internal blocking occurs because one directional coupler can not perform two operational states at the same time. This blocking is the main disadvantage of a 2x2 directional coupler switch.

A large scale optical crossbar switch is an integrated optic device which is usually fabricated on an electro-optic substrate such as lithium niobate ($LiNbO_3$) because this material is in the 3m crystallographic group and has a large linear electro-optic effect. It is vital to ion dope the substrate to cause impurities. The impure $LiNbO_3$ substrate induces light confinement in a channel of the waveguide.[31] Furthermore, the Ti-indiffused lithium niobate (Ti: $LiNbO_3$) directional couplers exhibit low optical loss, small operating voltage, and easy fabrication. Notice that the proceeding figures of the feed forward crossbar network topologies are understood to be integrated electro-optical devices, but will not be portrayed as such. However, they will be graphically illustrated as a combination of 2x2 switches. Hence, an integrated optic device is always an NxN switching array where N is an even number.

Although a variety of crossbar configurations has been proposed by various authors, a switch designer must consider a trade off between cost, complexity and performance. The banyan class of interconnection networks offers fast packet switching, the complexity is less than $O(N^2)$, and it has the potential to be a blockfree switch when used in conjunction with an appropriate sorter to alleviate node conflicts and output contention.[7,32] Another advantage of a banyan network is that it has a well defined architectural procedure to design the crossbar switch in any size. A large multistage switch network comprises smaller crossbar switches in the banyan network. A diagram of a 4x4 crossbar switch as shown in Fig. 1-5 is a configuration in which a banyan network will be built on for a larger switching array. The operational function of this switch is to allow connectivity among its input and output ports. For instance, input port 1 can be connected

to output port 4 by driving node A1 and B2 to the crossover state. If switch A1 and B1 are in the crossover state, it provides a path from input port 2 to output port 2, and input port 1 to either output port 3 or 4 depending upon the operational state of node B2. Similarly, if node A2 and B1 are in the straight-through state, a connection is established from input port 3 to output port 2, and input port 4 can be connected to either output port 3 or 4 depending upon an operational state of node B2.
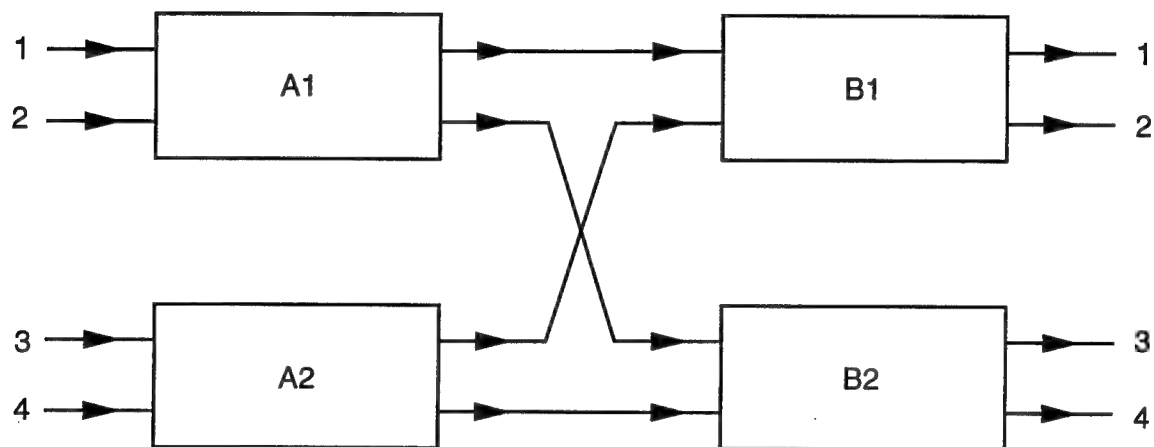
Fig. 1-5   A 4x4 feed forward interconnection network.

A switching network as shown in Fig. 1-6 is an 8x8 generic banyan network. Its architectural structure is simply a composition of two 4x4 crossbar networks lying parallel in the first two stages, while the last stage connections provide a unique path property for any interconnection permutations between all input-output pairs. There is no more than one path through the network for each input-to-output port interconnection. For instance, connectivity between input port 4 and output port 6 can be established by driving node A2 to crossover state, node B1 to straight-through state, and node C3 to crossover state. At the same time, other connections are also possible by following the directions of the arrow heads.

The name banyan was derived from the banyan tree because infrastructural interconnections of the network from stage to stage are similar to the branching of that tree. This network is often used as a standard for comparison with other interconnection networks, because it has only $1/2(N\log_2 N)$ complexities and is potentially a blockfree network when used in conjunction with a contention resolution network. Notice that normally an NXN input-to-output crossbar switching network requires N^2 hardware complexities.
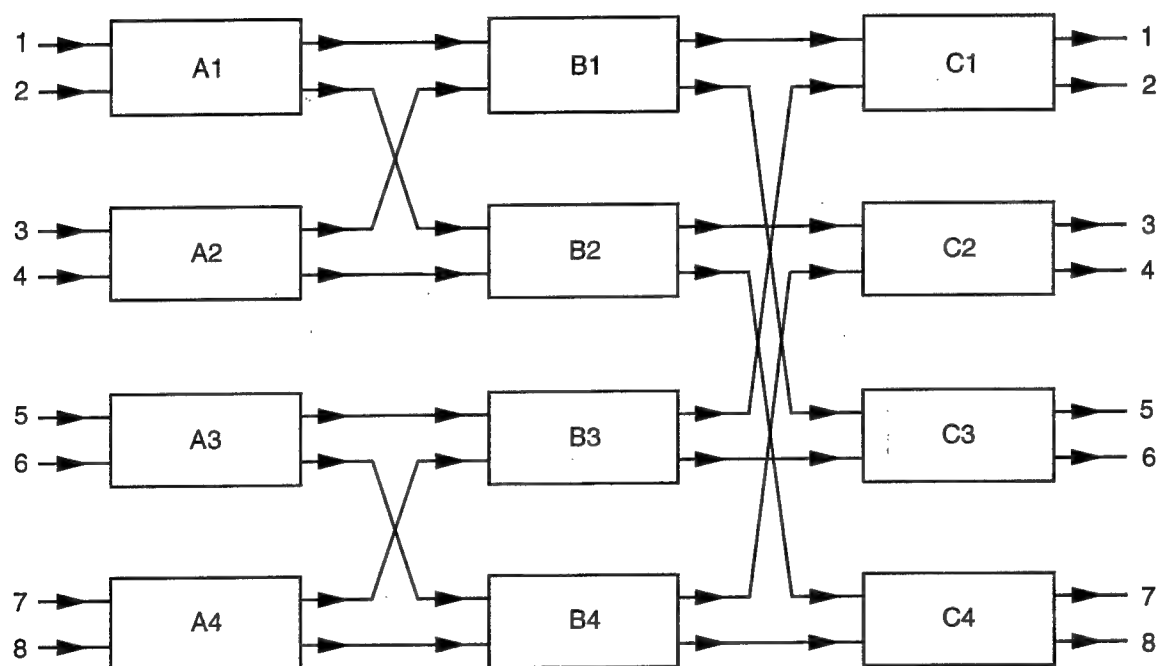
Fig. 1-6   An 8x8 banyan switching network.

Let's scale-up a banyan switching network to a 16x16 input-output array as shown in Fig. 1-7. The purpose is to graphically demonstrate how a large scale banyan network topology can be constructed as a composition of smaller ones. A 16x16 banyan network consists of two 8x8 banyan switching networks lying parallel in the first three stages, and connected together by the fourth stage to allow crossbar permutations. The banyan switch

is a multistage network topology, but an architectural procedure is well defined for any input-output array size construction. In other words, it is a scalable switching network.
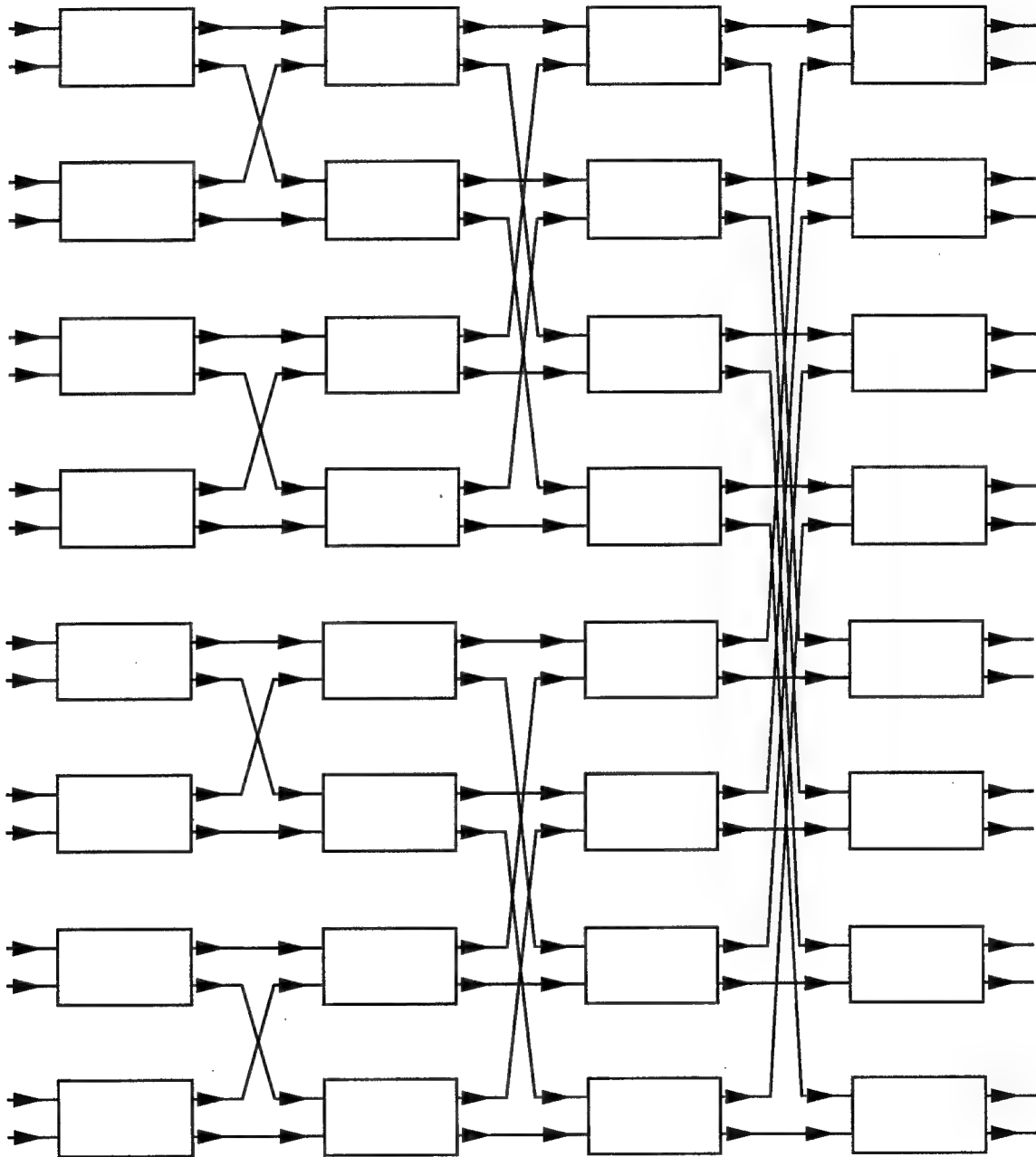
Fig. 1-7   A 16x16 banyan network topology.

1.6 Motivation for Free Space Optical Crossbar

The rate of hardware complexity in a guided optical crossbar configuration increases with the size of the input-output array. A guided crossbar switch is a multistage system. Each stage degrades the data being switched, ergo, insertion loss, crosstalk, and signal distortion are compounded from stage to stage. This infidelity impacts greatly on signal-to-noise ratio which could tremendously increase bit error rates as the size of the network is enlarged.[33] In theory, transmitted data can become unintelligible after it has navigated through a large scale switching network. One of the reasons is the loss in a directional coupler which is estimated to be 0.8 dB/bend and 3 dB/crossover for a very conservative design using current lithium niobate technology.[34] Hence, insertion loss alone can contribute a major design constraint on optical switches. In fact, a guided optical crossbar switch is only practical for a small scale network. Furthermore, a guided switching network must resolve node and output contention which can be eliminated with a sophisticated hardware design. However, a contemporary switch designer must consider a fundamental trade-off between hardware complexity and hardware cost, and implement an idea that provides optimum performance and minimum cost without sacrificing the integrity of the information being transmitted. These requirements have posed a very challenging task. Research and development has been conducted on various guided optical interconnection topologies. A universal objective is to achieve a nonblocking architecture.
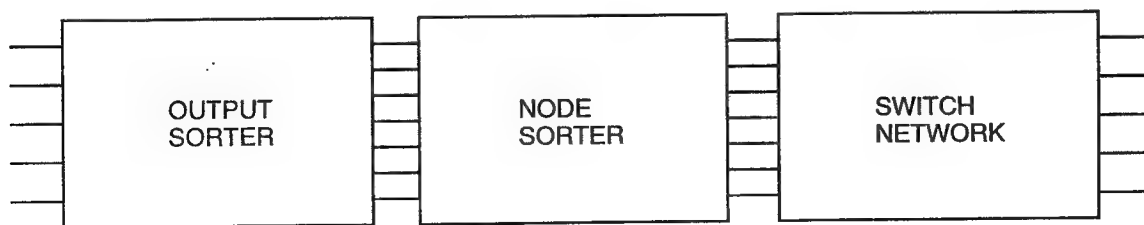
Fig. 1-8 Blockfree optical guided switching network topology.

24

One approach is to treat a switching system as three separate functional networks as shown in Fig. 1-8. It consists of a sort and switch network. The function of the sort network is to resolve node and output contentions. These sorters are called the node sorter and output sorter. The function of the switch network is to provide connectivities between input and output ports. A practical implementation of a nonblocking switch is realizable by a Batcher-Banyan network. The name derives from the fact that this network is a product of two separate components combined together. Banyan is the switch network responsible for providing input-output connectivities. This network has been described in detail in the previous sections.

The responsibility of a sort network is to resolve data traffic conflicts before they enter a switch network for routing processes. Batcher sorting networks are the subject of this discussion because it has a fast sorting capability and hardware complexity in the $O(N\log_2 N)$ which is in the same order as banyan network. This sorting network is also called a bitonic sorter which was credited to K. E. Batcher in his work to utilize bitonic sequences in sorting algorithms.[35] The bitonic sequence is two monotonic sequences, ascending and descending order, in juxtaposition. If a given sequence is bitonic then the sorted sequences are also bitonic. Let's consider a given bitonic sequence : $x_1, x_2, ..., x_{2k}$. The sorting algorithm can be determined by computing monotonically increasing and monotonically decreasing sequences. The two k-number sequences after bisecting the 2k-element sequence can be determined by minimum and maximum computations. The ascending sequence is $L_i = \min(x_i, x_{k+i})$ and the descending sequence is $R_i = \max(x_i, x_{k+i})$ for $1 \le i \le k$. Notice that every entry in the L_i sequence is less than every entry in the R_i sequence ($\{L_i\} < \{R_i\}$), where $\{L_i\}$ and $\{R_i\}$ is a left and right monotonic sequence, respectively. These sequences can be partitioned together to form a bitonic sequence in a Batcher sorter. By repeating this bitonic sorting process until each sequence is length 1, the entire given sequence becomes sorted. The final sorted sequence is in ascending order after it has been subjected to $\log_2(2k)$ sorting steps. The constraint of the bitonic sorter, k,

must be an even number, and a given sequence can not have repeated numbers. The requirement of an even k makes it suitable to employ 2x2 optical switches for a hardware implementation.

The hardware architecture of a large bitonic sorter can be constructed from a number of smaller ones. The smallest structure of a bitonic sorter is a 2x2 switch where its function is to compare the two incoming numbers and perform a conditional exchange to shift the smaller of the two to the output marked S and the larger one to the output marked L, as shown in Fig. 1-9.[36] A large scale bitonic sorter is a multistage system like a banyan network. In addition to a comparison-exchange function, a bitonic sorter also inherits a shuffle function which is responsible for the interconnection pattern between the nodes. The combination of these three functions makes it possible for a bitonic sorter to resolve node and output contentions. First, a bitonic sorter will sort out a header address to eliminate output contention when more than one packet is destined for the same port. The output sorter flags 0 to those conflict packets that have low priority, and flags 1 to one of the packets in a conflict set to signify a high priority when the packet is routed through the switch network.[35,37] Since the flag 1 packet takes precedence in a routing conflict, the flag 0 packets are subjected to recirculation through the switches where they are fed back to the inputs for a subsequent operation. Hence, the flag bit is the first bit of each packet with 0 signifying an inactive packet and 1 signifying an active packet. Packets exiting the output sorter are in a relative order as they reenter into a node sorter. These packets will be
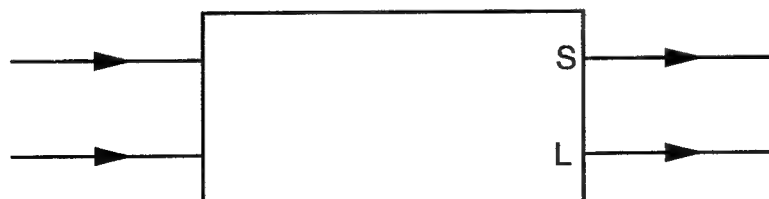
Fig. 1-9 A 2x2 comparison-exchange bitonic module.

shuffled around in the node sorter to avoid node collisions in the switching network. At the node sorter output, the packets have been shuffled and sorted into an absolute nonblocking order according to the packets' priorities and destination addresses. Finally, the switch network routes these nonblocked, sorted packets to appropriate output ports. The routing of packets through the switch network is called self-routing since the header of each packet contains all necessary destination information needed to navigate through the network. The incoming packets to the switch network could be asynchronous. However, the sorters require incoming packets to be synchronous for comparison operations. This introduces an extraneous complexity to both hardware and software to synchronize all packets before they enter a node sorter. A fiber optic feedback delay line can be used to synchronize data packets for asynchronous incoming packets. For instance, a lagged data packet can be switched to a feedback delay line to line up with the next set of data packets for the next sorting cycle.

The foundation has been laid to support the need for a high capacity free space switching fabric with dynamic holograms (HICAFDYH). The technical developments put forth in this dissertation will address a major technical improvement and advantages over a guided optical crossbar. The input-output connectivities in the HICAFDYH can be set up and torn down in real time. This switch is completely free of node contention since it is a free space system. Intelligence of the switch depends on the information of each address packet, ergo, the packet can be viewed as self-routing which has no restriction on synchronization because this system does not require an address comparison module. The system adopts a first-come, first-serve fashion, namely an asynchronous transfer mode (ATM), which is one of many advantages in the HICAFDYH. The traffic management is done by computer and data may not be sent without first establishing a connection. The explicit description, including system and node functionality, of the HICAFDYH will be reported in chapter 2. In addition, its application scenario is also provided in that chapter.

1.7   Concept of a Free Space Crossbar Switch

Interconnectivities of input-to-output mapping can be presented in a vector-matrix multiplication scheme as a novel concept of crossbar technologies. This multiplication is a mathematical expression for an interconnection system of a plural input link to each output port as shown in Fig. 1-10. The figure is an NxN holographic crossbar. A visual perception seems to be misled by this figure. It is not one input channel interconnected to all N output ports at the same time. Its intent and purpose is to illustrate a graphical
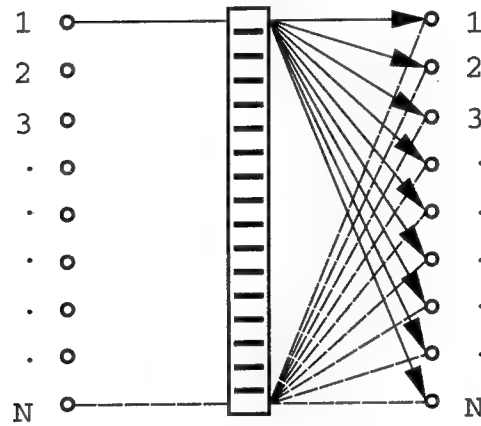
Fig. 1-10   A novel scheme of NxN holographic crossbar.

demonstration that each input port can be mapped to any output routing. Thus, an output port must anticipate output contention of having to receive multiple incoming signals at the same time. An output contention is one of the main problems in communication systems. The equation below will give some insight as to how this contention can be resolved. If y_i is assigned to represent an Nx1 output array, and input channels are denoted by x_i for Nx1 array matrix, where i = 1, 2, 3, ..., N, then an output matrix $[y_i]$ is a product of an NxN angular coefficient matrix to an input matrix $[x_i]$ as shown in Equ. (1-28) which is the matrix-vector multiplication equation.[38]

$$
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \cdot & \cdot & \cdot & \theta_{1N} \\ \theta_{21} & \theta_{22} & \theta_{23} & \cdot & \cdot & \cdot & \theta_{2N} \\ \theta_{31} & \theta_{32} & \theta_{33} & \cdot & \cdot & \cdot & \theta_{3N} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \theta_{N1} & \theta_{N2} & \theta_{N3} & \cdot & \cdot & \cdot & \theta_{NN} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} \qquad (1\text{-}28)
$$

where angle θ_{ij} is a weight coefficient. The coefficients are diffraction angles that provide interconnectivities from inputs j to outputs i, where $1 \leq (i, j) \leq N$. Employing both the graphical illustration of Fig. 1-10 and the matrix-vector multiplication of Equ. (1-28) will further elucidate a holographic crossbar concept to a new level. However, this system is not completely free of output contentions (many-to-one mapping). For a simplification, it is more suitable to express this equation in a summation form:

$$
y_i = \sum_{j=1}^{N} \theta_{ij} x_j, \text{ where } i = 1, 2, ..., N. \qquad (1\text{-}29)
$$

where output y_i is a dependent variable, input x_j is an independent variable, and subscript j varies from 1 to N (j = 1, 2, ..., N) for every ith index value. In general, a product $\theta_{ij}x_j$ is not only independent of $\theta_{ik}x_k$, but also they have an unequal weights.

Let's envision the arrival of destination requests at a node network in terms of data packet. Since packet arrivals are statistically independent random variables, it is reasonable to assume that there is always more than one packet having the same output request at the same time. It is unavoidable in a communications system to encounter a high probability of output contentions. But contention can be resolved by methods suggested in Equ. (1-29). One technique is to employ output conflict resolution that only permits routing passage to those requests with highest priority and denies access to those packets with lower priority

by mathematically setting their corresponding routing angular coefficients θ_{ij} to zero.

Another method requires a more sophisticated conflict resolution device such as a network that sorts out conflict packets by forward routing deflection. Instead of discriminately sanctioning certain packets for throughput, it deflects all incoming
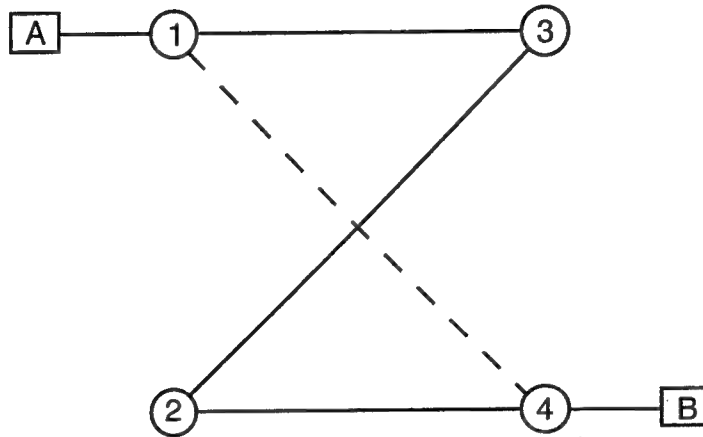
Fig. 1-11 Routing deflection throughputs of a communication network.

packets to alternative output channels in attempting to resolve output conflicts as shown in Fig. 1-11. Note that a physical switching process is done by simply writing a grating Λ that allows $\theta_{ij} = \theta_{kj}$, where $i \neq j$. A basic concept of this method is graphically illustrated by a solid line to show how network stations A and B communicate in a congested environment. For instance, to avoid conflict with a packet in a dashed line, alternative links are established via nodes 1, 3 and 4, or nodes 1, 3, 2 and 4. If two packets depart from node 1 at the same time, a packet traveling through the solid line will arrive at node 4 lagging behind the packet traveling through the dashed line. The dashed line shows a direct route from node 1 to node 4. This route is preserved for highest priority packets or is used in a conflict-free condition at the output of node 1. These output resolution methods are applicable to both circuit switching networks and packet switching networks.

1.8 Summary

This chapter began with a historical background of a crossbar switch technology and optical fiber transmission. The development of a directional coupler makes it possible to design a guided photonic crossbar. A key difference of an optical approach in comparison to its electronic counterpart is the routing data redistribution which preserves fiber optic bandwidth capacity and resolves optical-to-electronic and electronic-to-optical conversion bottlenecks. But the crossbar is a multistage architecture. Hence, signal degradation is compounded over the stages. Moreover, hardware complexity increases with the increasing of an input-output array size. Furthermore, internal blocking is inherent in a guided crossbar structure. Thus, a node sorter must be employed prior to the switching network. Therefore, a realistic application of this crossbar scheme is in a small scale network. However, a free-space optical crossbar can be scaled to a large interconnectivity density. It is a monostage system which is completely free of internal blocking. As a matter of fact, system and node functionalities of a high capacity free-space switching fabric with dynamic holograms is discussed in the next chapter. Hence, it is proper to conclude this chapter with the concept of a free-space crossbar, which provides a smooth transition between the two chapters.

Chapter 2

High Capacity Free Space Switching Fabric with Dynamic Holograms

The advantages of a free space crossbar over a guided version, and its electronic counterpart, include large interconnectivity density, transparent data redistribution, fiber optic bandwidth capacity, freedom from internal blocking (which is one of the major drawbacks of guided optical crossbars), minimizing physical size, the solution of data infidelity problems due to multistage guided systems that contribute to degradation of signal-to-noise ratio and increasing bit error rates, and eliminating electrical/optical conversion. The system takes advantage of the parallelism and multidimensionality inherent in optics and can be scaled to a large capacity (NxN) while maintaining low weight and portability which are projected requirements for future broadband communications.

The purpose of this chapter is to describe the HICAFDYH in full detail. This chapter is divided into six sections. It begins by attempting to alleviate output contentions. The proposed approach to resolve the contentions is called an optical routing deflection network which is offered in Sec. 2.1. The virtual circuit packet switching concept is illustrated in Sec. 2.2. An application scenario of voice/video/data communication is provided in Sec. 2.3. System and node functionalities of the HICAFDYH are discussed in Sec. 2.4 and Sec. 2.5, respectively. The last section (Sec. 2.6) belongs to the routing controller.

2.1 Optical Routing Deflection Network

When multiple incoming packets are destined for the same output, only a packet that has the highest priority is allowed to have direct access. The rest of the packets must either be dropped, queued, or deflected. Dropping packets will require retransmission which wastes capacity and hinders performance. Queuing can occur at switch inputs or switch outputs. Queuing is a situation where data is forced to wait. Hence, data must be buffered into a delay line. However, queue size could grow without bound if the packet arrival rate exceeds the transmission rate. An alternative for traffic control is a deflection routing technique which can be used to control traffic congestion in a virtual circuit application.

Figure. 2-1 illustrates the concept of a deflection switching network which can be employed as a tool to resolve output contention. This network consists of collective interconnection nodes that provide access for data transfer from source to sink. The purpose of each node is not to be concerned with the content of data being routed, but to
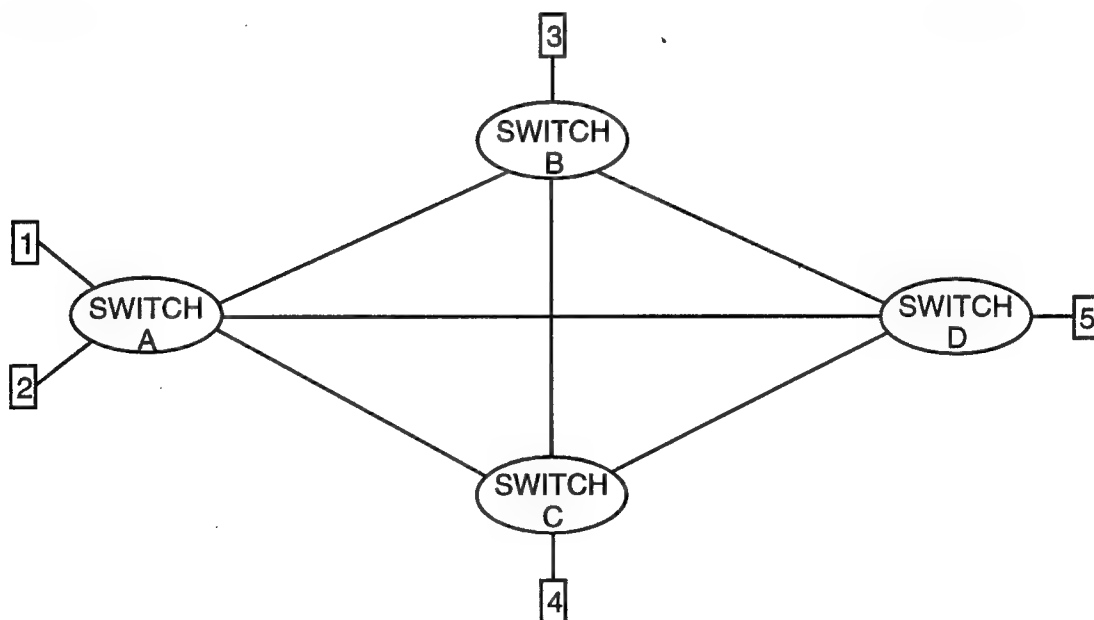
Fig. 2-1 Illustration of output resolution network by mesh topology.

provide a facility to navigate data from node to node until it reaches its destination. The network shown consists of five stations, 1 through 5, attached to four switches, A through D. Station refers to a communication device, such as telephone or computer. Each node is controlled by an intelligent traffic control switching system which requires a software intense operation (high level function) to resolve output contentions. For example, station 1 can communicate to station 5 via switches A and D; switches A, B and D; and switches A, C and D. Therefore, switch A must decide on a link that has the least traffic congestion. Notice that if the link between switches B and D fails, then switch B reconfigures a new path to destination 5 via switch C. Similarly, switch C can reroute a link to destination 5 via switch B if the connection between C and D fails. The example has demonstrated three possible combinations in which station 1 or 2 can communicate with station 5. In general, the mesh topology, as shown in the figure, provides (N-1) possible output combinations, where N is a number of total nodes. Hence, the probability that each outgoing link will be selected is 1/n where n is the total outgoing links attached to the decision node (n = N-1). Since each outgoing link has equal probability, which is a uniform distribution function, this scheme may not provide good traffic distribution. Alternatively, the probability could also be based on data distribution per link. Therefore, the probability that link j will not be selected is :

$$P_j = \frac{D_j}{\sum\limits_{i=1}^{n} D_i} \tag{2-1}$$

where P_j is the probability of not selecting outgoing link j, D_j is data distribution on link j, and n is the total number of outgoing links attached to the decision node. The sum is taken over all outgoing links. Thus, the probability that link j will be selected is :

$$Q_j = 1 - P_j \tag{2-2}$$

There are two types of blocking at each node - internal blocking and output blocking. A free-space crossbar is inherently free of internal blocking, but contentions can

still occur at the outputs due to k-to-1 mapping. An example of output contention is when incoming data from two different ports request the same output port. In order to optimize performance of output conflict resolution, one must ask how often a node receives multiple requests for the same destination. The probability could be based on input/output links. Let's consider that each node has n inputs and n outputs attached to it. The probability that at least two requests are destined for the same output port is one minus the probability that no two requests have the same destination. The functional expression of this probability becomes :

$$P(x \geq 2) = 1 - \frac{n!}{n^n} \qquad (2\text{-}3)$$

where x is the number of conflicting requests on a single outgoing link. Notice that this equation is independent of data arrival rate (bursty or not) which, in a classical sense, is a Poisson distribution. The probability quickly approaches unity as the number of crossbars increase as shown in Fig 2-2. Hence, the output resolution process may require both methods - low priority elimination and a sort system by routing deflection, to resolve output contention in some cases.
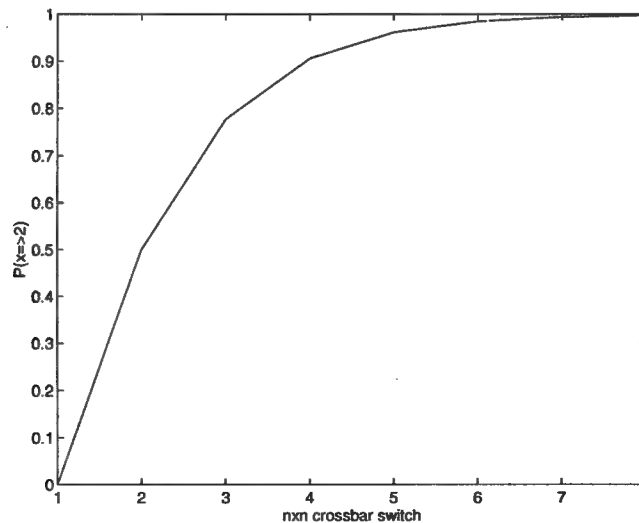
Fig. 2-2 Output contention probability versus crossbar dimension.

2.2   Virtual Circuit Packet Switching

Virtual circuit packet switching (VCPS) is a hybrid between circuit switching and packet switching. It adopts both circuit establishment and message transmission in a packet form. VCPS requires a dedicated path through a sequence of nodes between two communicating stations. A request packet (RPK) must first be sent to establish the path by connecting node to node until it reaches the destination station. The received station then sends an acknowledgment packet (APK) back to the requesting station to initiate the data transmission. The data is transmitted in a packet form. Thus, the transmission process appears quite similar to packet switching at this juncture. Note that after a dedicated path has been established, data redistributions are effectively transparent through each node. Hence, node delay only occurs during circuit establishment as shown in Fig. 2-3 (which was extracted from reference [1]). The purpose of this illustration is to graphically demonstrate the basic principle of VCPS data transmission.

The communication switching scheme illustrated in Fig. 2-3a consists of three nodes (N = 3), with at least two stations attached to them. It shows that a communication path goes through a series of nodes 1, 2 and 3. The connection is initiated by Node 1. It sends a deterministic destination packet, which is a RPK, via node 2 to node 3. Node 2 must make a decision once it has received the packet. It has to decide whether to grant access for continuing circuit establishment to the next node, or to reply with a negative acknowledgment to deny access if a traffic is too congested. If there is no contention at the node-2 outputs, then circuit establishment will continue to node 3. Node 3 searches out the status of a destination station. If there is no conflict then a positive acknowledgment packet (PAK) will be replied to node 1, confirming that a dedicated path is established between the two stations that are attaching to nodes 1 and 3. Now, these two stations can begin data communication. Data packets (PKTs), as shown in the figure, were labeled as PKT1, PKT2, PKT3, etc. These packets do not experience node delay during the entire transmission. However, they are subjected to propagation delay which is the time it takes

36

for a packet to propagate from node 1 to node 3. The time it takes for all packets to transmit from source to destination including propagation delay is represented by t_{PK}. Note that the RPK and APK packet sizes are much smaller than the size of PKT. Hence, one application of this virtual switching technique is in the area of a full-duplex communication system - live videophone.
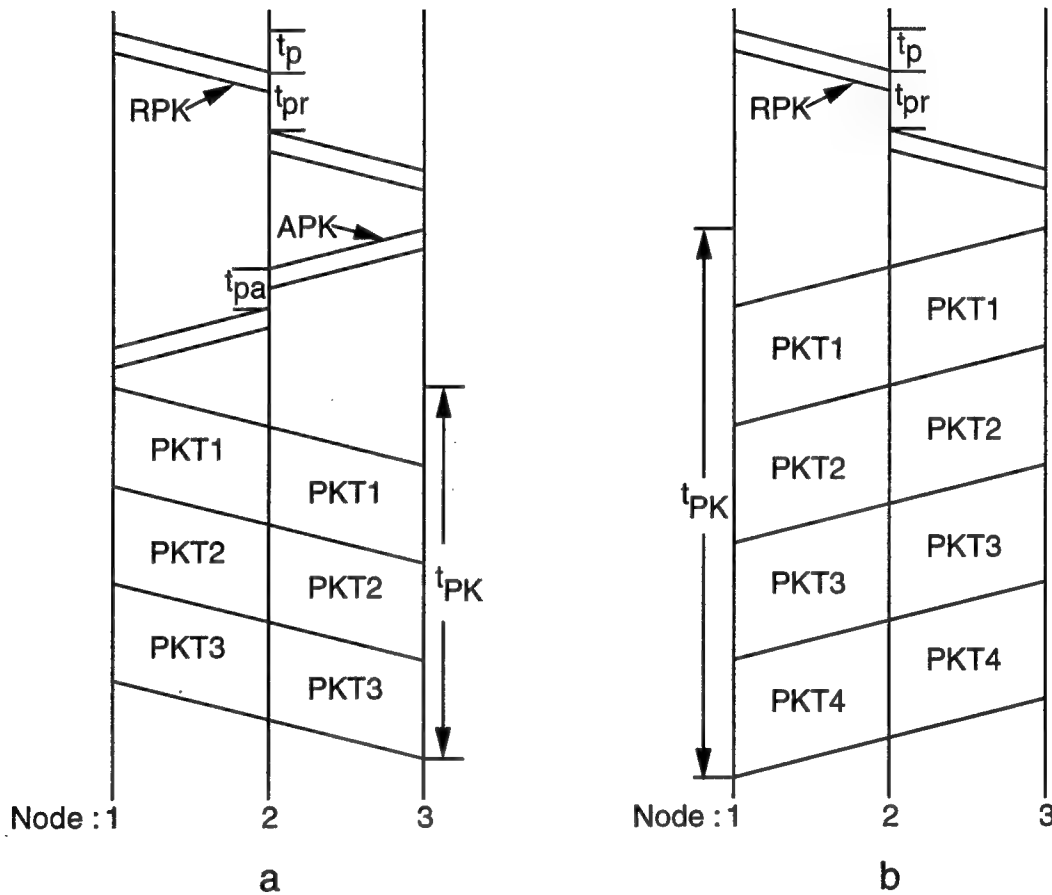
Fig. 2-3 Virtual circuit packet switching scheme.

Applications : a.) Voice/video/data communications b.) Video/message-on-demand.

During circuit establishment, RPK is subjected to propagation delay t_p in addition to processing delay t_{pr}. Processing delay is the time that each node requires to set up the routing interconnection. A replied APK also undergoes node delay (t_{pa}), due to node-to-node retransmission process, as well as propagation delay t_p. Note that t_{pa} is generally

much less than t_{pr}. If t_a represents the total elapsed time that is needed to establish a dedicated path, the functional relationship is expressed as :

$$t_a = \sum_{i=1}^{N-1} 2t_{pi} + t_{pri} + t_{pai} \tag{2-4}$$

One application of this switching scheme, as shown in Fig. 2-3b, is in the area of half-duplex data transmission, such as video-on-demand or message-on-demand, where a request must be initiated by a customer via a series of nodes to retrieve information. The function of each node is either to forward request packets to a data distribution center that controls central archives, or to reply with a negative acknowledgment back to the customer to repeat the request after some preset time. A request packet not only experiences propagation delay from source to destination but also undergoes processing delay through each node. However, a circuit is fully established as soon as the packet reaches the data distribution center, which in turn, promptly transmits the requested data. Hence, the total elapsed time (t_b) for circuit establishment is :

$$t_b = \sum_{i=1}^{N-1} t_{pi} + t_{pri} \tag{2-5}$$

The relation between t_a and t_b depends on transmission media and data processing methods that are employed to compute switch permutations. Thus, the relation of these two elapsed times may not have any significant meaning since each switching technique has its own application.

The performance of the virtual circuit packet switch can be evaluated by studying its maximum potential efficiency or utilization[1] (U) of each source-destination connection. It is defined as

$$U = \frac{t_{pk}}{t_i + t_{pk}} \; ; \quad \textbf{where} \quad \textbf{i = a, b} \tag{2-6}$$

If t_{pk} is much less than t_i ($t_{pk} \ll t_i$) then U is a number much less than one, in which case VCPS is not a suitable candidate. This condition requires a high speed switching

38

network, namely, packet switching. The speed of a packet switch, including process routing information and the setting of switching elements, must be faster than the packet arrival rate in order to avoid a data flow bottleneck at the switch input.[39] On the other hand, if t_{Pk} is much greater than t_i ($t_{Pk} \gg t_i$), then U approaches unity. In this case, VCPS is a very efficient technique. As a matter of fact, VCPS scheme is one of the most qualified candidates in a long data stream communication. The proceeding section is devoted to one of the examples chosen to illustrate VCPS application.

2.3 Voice/Video/Data Communication Network

The installation of optical fiber has grown explosively within buildings, across cities, states, countries, and continents. However, the switching technology has not fully capitalized on this fiberoptic infrastructure. The system that is put forth in Fig. 2-4 will attempt to take advantage of the bandwidth that fiber has to offer. This system responds to the challenge of a modern voice/video/data communication requirement. It can handle multiple data formats including voice, video and data.[40] This tremendous bandwidth will facilitate high-level processing functions because it can provide the luxury for high resolution transmission and high data rate. For example, this immense transmission capacity makes it possible to transmit moving video images in real time without resorting to data compression techniques.[41] The applications like room-to-room videoconferencing services, desktop conferencing services, and videophone have an insatiable hunger for bandwidth.[42] This switching system will meet those demands. Furthermore, it can bridge conferences originating from two or more locations by multiple-order diffraction techniques.

Figure 2-4 depicts the network configuration for voice/video/data communications in a wide area network. The system was simplified to four nodes to demonstrate a basic
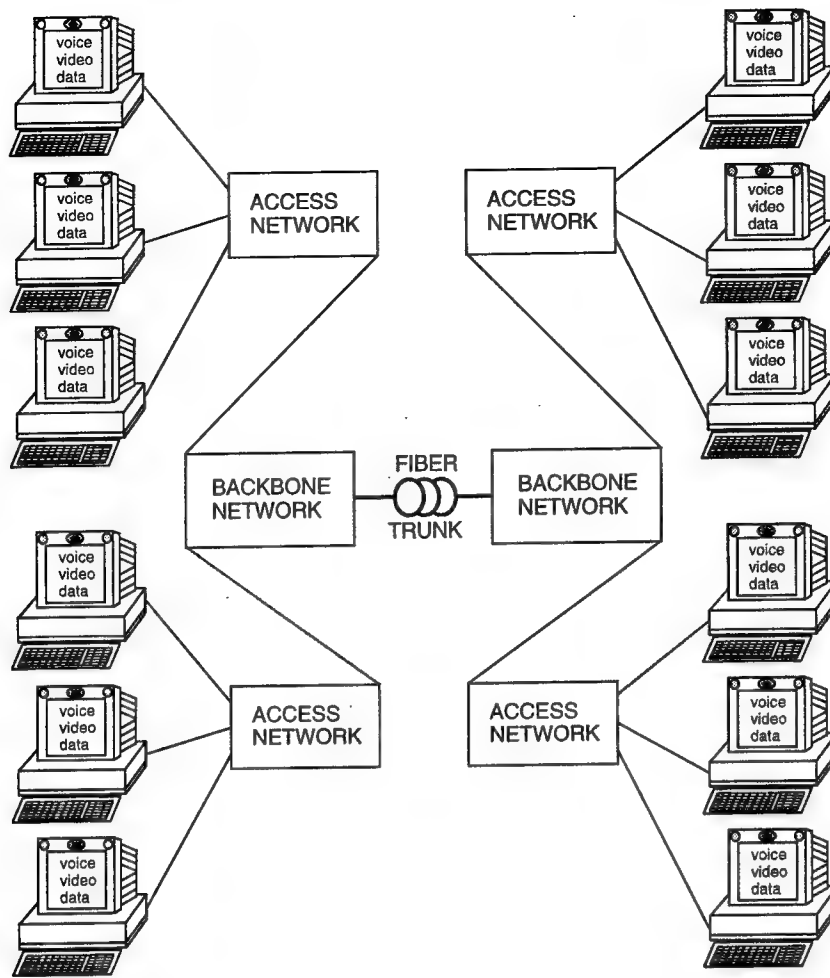
Fig. 2-4 Network configuration for voice/video/data communications.

videoconferencing concept. A cluster of users connects to an access network, and a number of access networks links users together via a series of backbone networks. If the dimension of an access network is NxN (N inputs and N outputs), then a backbone network has dimension of ANxAN, where A is a number of access networks connecting to a backbone network. In spite of their different dimensions, both networks, access and backbone, have the same switching functionalities. The switching system of these networks are free-space, dynamically reconfigurable, bacteriorhodopsin, holographic crossbars. The details and functionality of this crossbar will be described in the following sections.

Let's say a user x on the left hand side (xL) of the figure wants to communicate with a user y on the right hand side (yR). User xL must first try to establish a communication link with user yR by sending a request packet to the access network. The request is then forwarded to user yR via two backbone networks and an access network which locally connects to user yR. If yR is not busy, it replies with a PAK to xL signifying the completion of a circuit establishment, or else it returns a negative acknowledgment (NAK) to xL denying the request. Likewise, each node can also return NAK to a sender if it is experiencing a contention. In addition, a sender will have to resubmit a request after a certain preset time has expired if it has not yet received any message from a receiver or network node. Note that both packets, request and acknowledgment, are subjected to processing delay at each node as shown in Fig. 2-3a. The foundation of this concept is to graphically illustrate a full duplex system.

2.4 Proposed System

An optical crossbar switching system, as shown in Fig. 2-5, is a device that allows data to travel from N input ports to N output ports in the optical domain. This system requires only request/acknowledgment (address packet) to be in a packet form. It is a partition of flag, error check information, priority, address length and destination address. The arrival of these packets at each switching node is assumed to be random. Hence, the routing controller processes information on a first-in first-out (FIFO) manner. In general, packets in N parallel pipelines arrive at each input node asynchronously, but switch permutations of input-output mappings are not truly an independent function. For example, a switch permutation of many-to-one mapping (output conflict) from input ports to an output port requires that the highest priority packet be routed correctly while choices are made on the remaining packets to deflect , drop or buffer. The priority bits in a request packet become very important if and only if two or more packets arrive at a node input

41

simultaneously. In this case, a correct routing is yielded to the packet with a highest priority.

This crossbar switch is used in concert with a routing controller illustrated in the figure. The switch first sends the routing address to the controller for a mapping permutation process. It then dispatches routing state control information to the switch to initiate a physical routing set-up. At the same time, it also forwards address packets to the next logical node. This crossbar system is a hierarchical structure. The crossbar switch belongs to a physical link layer in the OSI model (open systems interconnection) which was developed by the International Organization for Standardization. The routing controller is part of a higher functional layer of the hierarchy. The functionalities of these two layers will be explored in a full detail in the next two subsections.
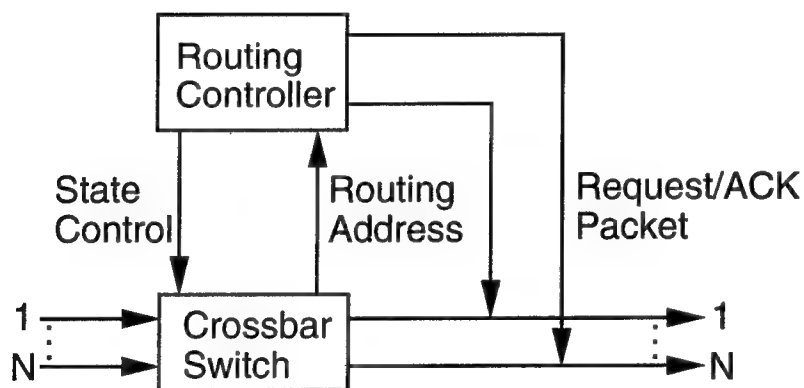
Fig. 2.5 Free-space dynamically reconfigurable optical crossbar system.

2.5 Node Implementation

Description of a free-space node functionality, as shown in Fig. 2-6, for a novel bacteriorhodopsin (BR) based photonic crossbar system for broadband communications is reported.[43, 44] A single beam from a laser source is collimated by a beam expander to fully

illuminate a binary spatial light modulator (SLM) which has dimensions of 256 by 256 pixels. Semetex offers an electronically controlled high resolution, high speed 256 x 256 SLM array. Its pixel spacing is 60.0 μm, and its performance is 125 frames/second (8 msec./frame). This SLM can handle 256 input-output mappings, and each mapping has 128 switch permutations. Utilizing two symmetric pixels with respect to an arbitrary center forms a Bragg angle in the BR film, in which a corresponding grating period is written.
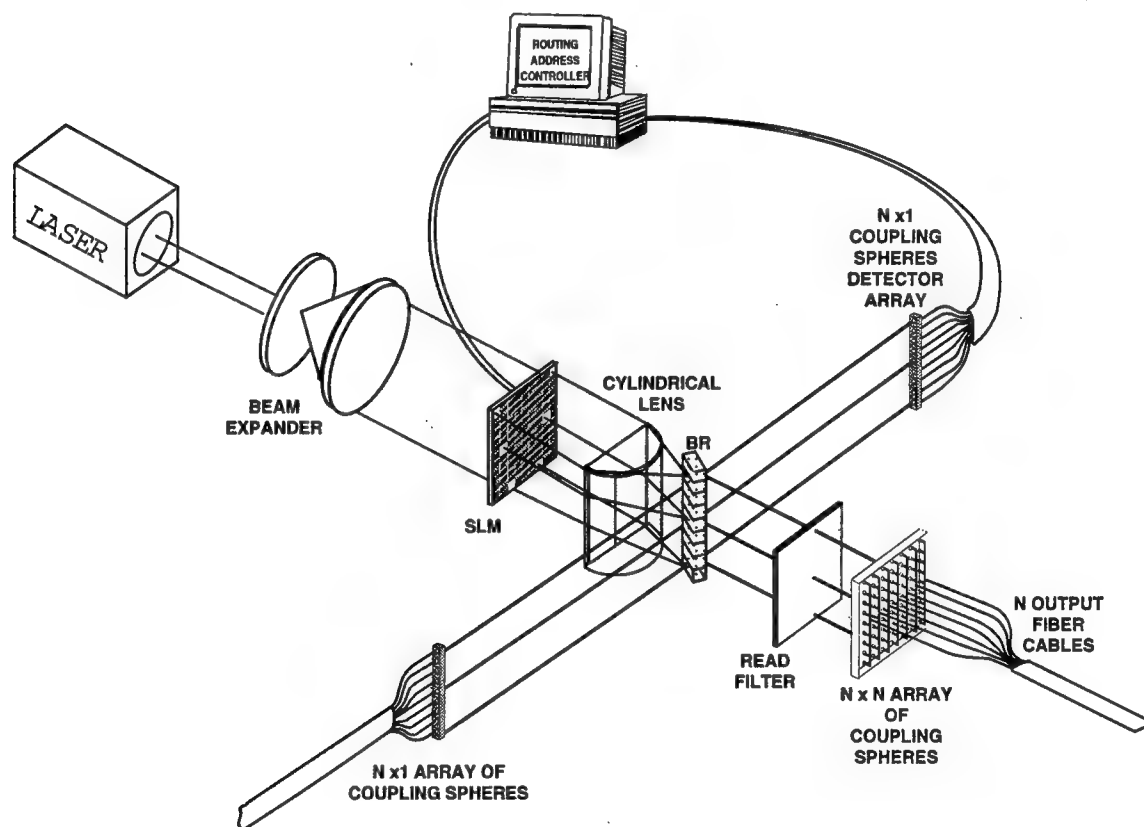
Fig. 2-6 Free-space dynamically reconfigurable bacteriorhodopsin crossbar.

These pixels must be in the same row, and beamlets are focused onto the film by a cylindrical lens. This SLM is electrically addressed by a routing address controller after it has extracted routing information from the address packets. A zeroth order diffraction

beam carries the destination address information from the transmitter to the detector array which connects to the controller. Note that these packets must be sent prior to data transmission. An Nx1 array of spherical balls was placed at the focal distance away from an array of Nx1 fiber ends to collimate the throughput signals. The balls were also placed at the focal distance in front of a detector array that is connected to the controller. After holographic gratings have been written in the film, read signals from the input array are diffracted to their respective destination channels. The function of the read filter in front of the output array is to block out beams that are not part of the read signals. Note that Nx1 input ports should have Nx1 corresponding output channels. There are NxN coupling spheres in the output array. These spherical balls were placed at the focal distance away from N-star couplers, where each has N-to-1 composition as shown in Fig. 2-7. This arrangement is necessary to accommodate all N input links. The concept is to map an nth row in an Nx1 input network to any port in an nth row of an output network depending on the destination request, and any signal from an ith column of the network is fed to an ith corresponding output channel. Again, the NxN array of spherical balls and the NxN array of fiber ends are separated by the focal distance.
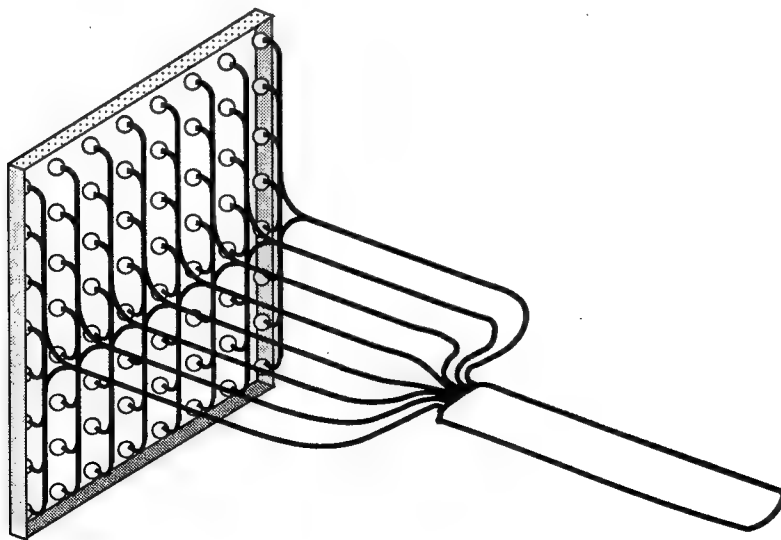
Fig. 2-7 N-output network with NxN coupling spherical balls.

The use of fiber coupling spheres is one of the techniques that is presently available for fiber beam collimation. Placing a spherical ball at the focal distance away from a fiber end produces a collimated beam at the output of the sphere. Inversely, the sphere focuses a collimated beam at the focal length plane. The application of two spheres in axial alignment, as shown in Fig. 2-8, transfers the focal point from one fiber end to the other. Hence, they can be used to couple fibers to fibers or fibers to detectors, etc. The advantages of coupling spheres are the ease of alignment and manufacture compared to thin lenses of small diameter.
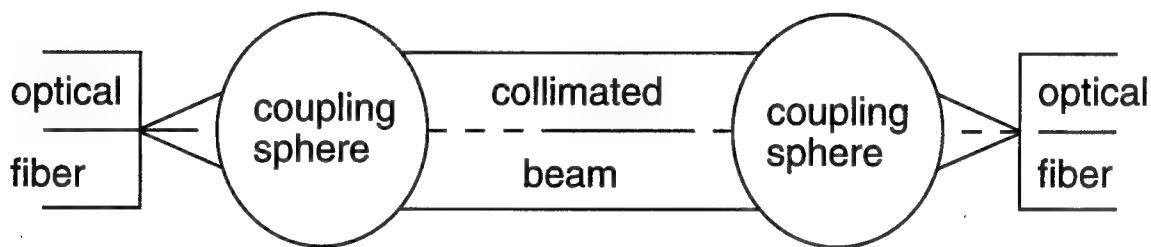
Fig. 2-8   Fiber-to-fiber coupling by spherical balls.

Melles Griot offers the spheres with three different diameters: 2.0, 3.0 and 5.0 mm with ± 1μm precision. Specifications claim to have only 0.4 dB insertion loss in a fiber-to-fiber coupling application as shown in the figure. Furthermore, off-axis aberrations are not a main concern because the spherical focal length is much larger than the fiber diameter.

2.6   Routing Controller

The address packet was briefly mentioned in previous sections. Let's organize this packet into frame format, as shown in Fig. 2-9, to illustrate its purpose. This frame is a

partition of flag field (FLG), error-detecting code (EC), priority bits (PRT), address length (LGH), and address field (ADD). The pattern in a flag field should be unique, and in reference [1], this pattern consists of 01111110. Of course, there is no assurance that this 8-bit pattern ever occurs in a data stream. This problem can be rectified by employing a bit stuffing procedure. This procedure requires a transmitter to always insert an extra 0 bit after each consecutive occurrence of five 1's in a data stream, with this zero bit eventually removed by the receiver. The flag field is a preamble bit pattern to indicate the beginning of an address packet. The information about error detection and correction are coded in EC partition. The priority information of each packet falls in PRT. The identification of the packet size and address length are coded in LGH. Its purpose is to be used as a flag for the postamble to strip the destination address from the packet, so it can be used in nodal source-destination processing. An address field is the portion of data frame that contains actual information about the routing destination.

FLAG	EC	PRT	LGH	ADDRESS
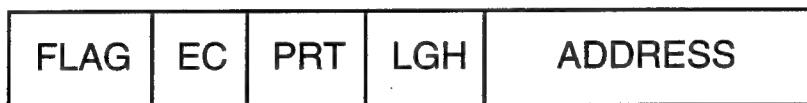
Fig. 2-9   Frame format of an address packet.

The decision to use an optical processor or an electronic processor depends on two factors: cost and processing speed. Electronic technology has reached its maturity. Hence, cost of an electronic processor is much lower than its counterpart, the optical processor, which is still in its infancy state. Furthermore, electronic processors can perform a more complex and sophisticated function compared to optical processors which can only implement a relatively simple processing function. Nevertheless, optical processors not only offers processing speed but also parallelism and multidimensionality.[39] This suggests that incorporating both processors in a design will capture the full advantages

46

that both worlds have to offer. A suitable application of electronic processors is in high-level functions, such as session setup. For a low-level function with a simple processing requirement it is best to employ optical processors. Let's suspend a decision on what kind of a processor will offer maximum potential. A decision to use optical or electronic processors should not be a burden at this junction. However, tasks and functionalities of
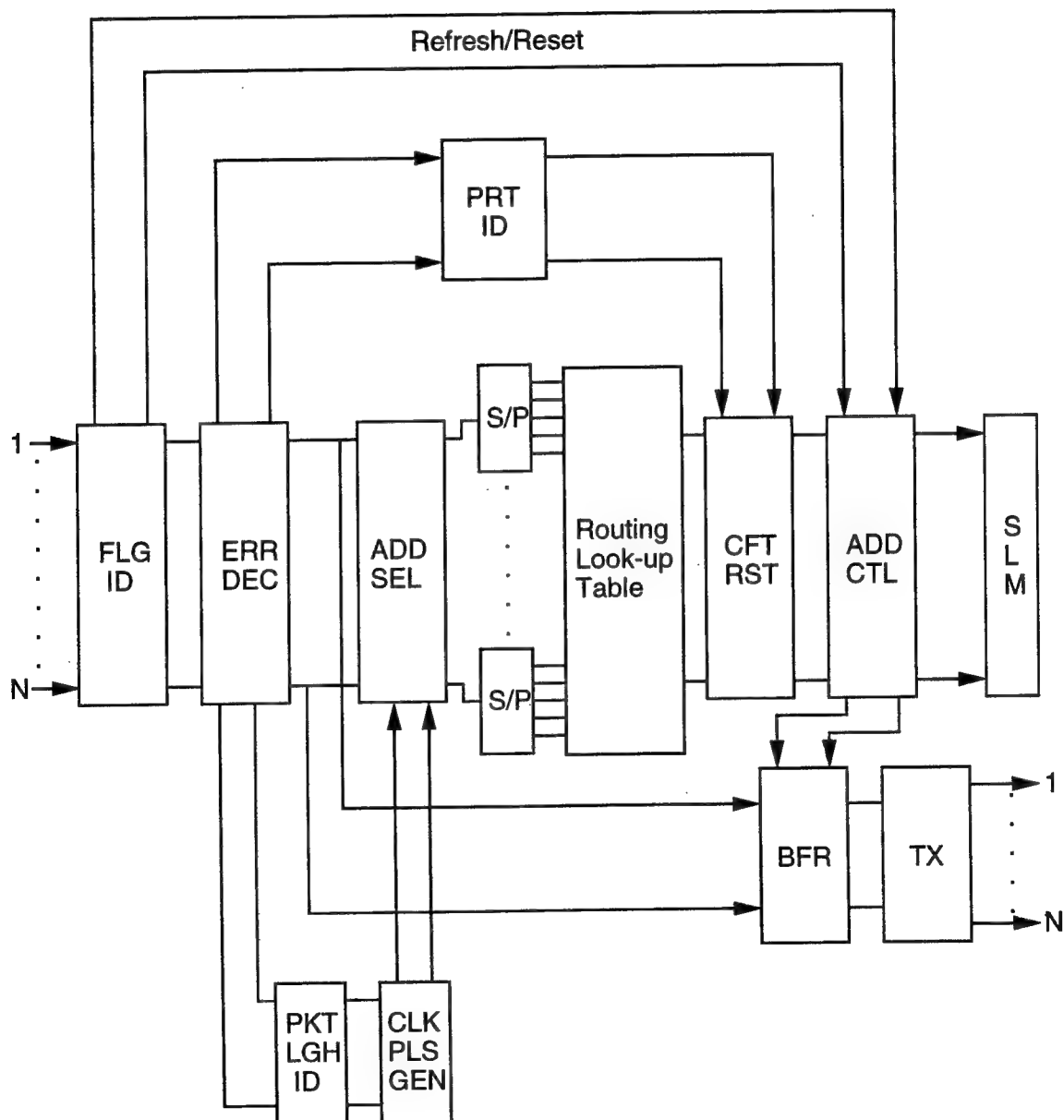
Fig 2-10 Block diagram of a crossbar switch routing controller.

each processor, as shown in Fig. 2-10, must be described clearly in order to fully understand the operational functions of this routing controller.

All processors used in the figure have N fan-ins and N fan-outs. It is called a parallel processor where all N packets can be processed simultaneously. The flag identifier processor (FLG ID) is used to identify the preamble bit pattern of each incoming address packet. If this pattern is identified, FLG ID forwards the rest of the packet to ERR DEC where a subsequent control field will be processed. Otherwise, the FLG ID sends a signal to ADD CTL to reset the sample by erasing the grating period with a single beam illumination if it does not detect any data flow. If FLG ID detects a stream of binary bits, but it is not a predeterministic preamble bit pattern, then it will instruct ADD CTL to refresh a sample periodically.

The function of the error decoder (ERR DEC) is to detect and correct errors that have occurred during data transmission. It decodes an error-detection scheme of each received packet and performs the same calculation as was coded into the packet by the transmitter. A received packet has error (s) if the two results are mismatched. This error-detection technique is known as a cyclic redundancy check (CRC). Thus, a decoder will attempt to execute correction procedures according to an error-correcting code if an error is detected in a packet. This technique is referred to as forward error correction because the error is corrected by a decoder, as opposed to backward error correction which is an application of retransmission schemes. Finally, sanitized packets are piped off to PRT ID, ADD SEL, PKT LGH ID and BFR, and each of these processors commences processing the packets according to their assigned functions.

A priority identifier (PRT ID) is used to identify the gravity of each packet, which was coded into PRT by the transmitter. The priority information is sent to CFT RST by PRT ID after it is decoded. CFT RST resolves output conflicts according to each packet's priority. For example, a direct routing is reserved for a packet with the highest priority to capitalize on the shortest transmission distance and time. Furthermore, a high priority

48

packet has first claim to this direct and dedicated path. A packet with a lower priority is deflected around to an alternate routing, and it has no assurance that it will reach its destination. Note that information about priority only becomes important in the case of many-to-one mapping from input to output interconnectivities, otherwise this information is discarded by the CFT RST.

The packet length identifier (PKT LGH ID) decodes the LGH partition of each address packet to determine the beginning and the end of each address field. After the address length is identified, the clock pulse generator (CLK) sends a synchronized pulse that has the same duration as the entire address field to the address selector (ADD SEL). The ADD SEL strips the address field from the rest of the packet by a vector multiplication operation as shown in Fig. 2-11. Note that a generated pulse must be perfectly synchronized with an address field for the correct address identification. At this junction, ADD SEL discards the control field from the packet, and it passes the routing address to the serial-to-parallel (S/P) processor.
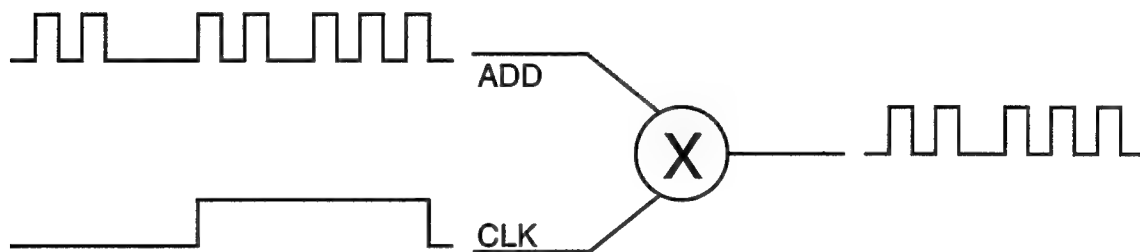
Fig. 2-11 Operational function of an address selector.

The function of the S/P processor is to transform a temporal pulse train into a spatial pattern.[45, 46] One way to achieve this transformation is to utilize an optical serial-to-parallel converter. A skewed pulse train after diffraction from a reflection grating can be transformed from a temporal profile into a spatial pattern with a symmetric bi-convex lens. This Fourier transform decomposes a function in part into a phase factor. This phase factor

can be eliminated by nonlinear wave-mixing with a corresponding phase conjugate wave. After phase cancellation, the pattern becomes time independent. Using a second symmetric bi-convex lens transforms this pattern into a replicated pattern in space corresponding to the input-pulse train. Finally, the S/P processor fans out this parallel address field to the next processor.

A routing look-up table is a processor that correlates a routing address field with a deterministic destination table which is stored in a look-up memory. This processor can achieve a K-fold increase in speed compared to serial performance since the correlation of a routing address field to the addresses in the table is performed in parallel, where K is the bit number of an address field. One advantage of a deterministic address look-up table is each address has been preset to a corresponding output port. Since each address occupies a specific location in the table, the physical location of each address defines a diffraction angle which corresponds to a specific output channel. Note that this processor maps a signal with corresponding diffraction information from an ith fan-in port to an ith fan-out port. The concept of this correlation is to have the signal represented by a high value at the output of the processor if an entire address field is exactly matched to one of those addresses in the table. Hence, the operational function of this correlator is very similar to an Exclusive NOR logic operator which produces output 1 when all inputs are in the same state as shown in the truth table in Fig. 2-12.[47, 48] A truth table of a two-input device can be described according to a Boolean functional equation as follows:

$$Z_i = \overline{X}_i \bullet \overline{Y}_i + X_i \bullet Y_i \; ; \; \text{where} \; i = 1, \; 2, \; . \; . \; ., \; K. \qquad (2\text{-}7)$$

Let X_i be a routing address field, Y_i a signal from the table, and Z_i an Exclusive NOR function. If all Z_i's are asserted high (all 1's), then and only then, the match is found, and the processor sends routing information to the Conflict resolution (CFT RST) processor.

The CFT RST processor is used to monitor output contentions. If there is more than one request destined for a single output port, the processor will resolve this conflict

according to the priority information from the PRT ID. It reserves a direct path for the packet with the highest priority, and deflects those packets with lower priority to alternate routes. In case of one-to-one mapping (no contention), this processor is inactive, and signals from routing look-up table flows through it to the address control (ADD CTL) processor. Note that all conflicts are resolved before signal has reached the ADD CTL. The ADD CTL processor electrically addresses a SLM which turns on a combination of any

X_i	Y_i	Z_i
0	0	1
0	1	0
1	0	0
1	1	1
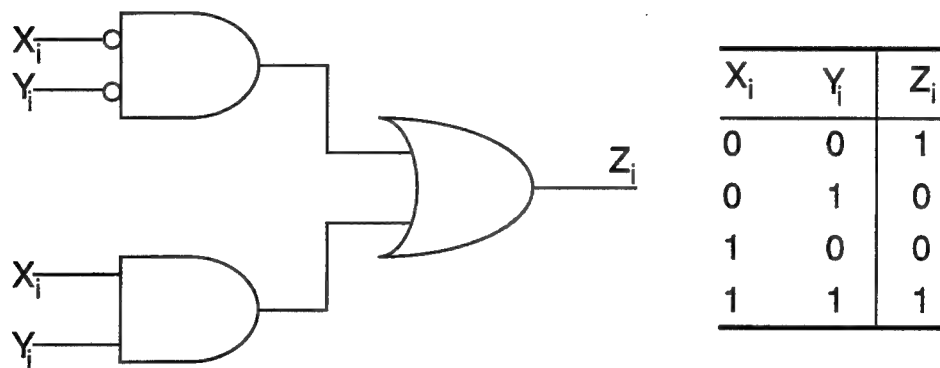
Fig. 2-12 Functional diagram for an Exclusive NOR logic and its truth table.

two pixels from the same row to transparent mode. A cylindrical lens focuses these two beamlets onto a photorefractive sample where gratings will be written. During a switch setup, which includes the SLM's frame switching time and decay/rise time of the sample, ADD CTL signals the buffer (BFR) to relinquish the corresponding address packet to the transmitter (TX). Note that this duplicated packet is retransmitted to the next node before the switch setup is complete. This decentralized switching process, where each node is required to retransmit the address packet to establish the link, can save the circuit establishment time because all nodes in the circuit can virtually process the switching access code and implement interconnection setup almost at the same time. Furthermore, the propagation delay time is generally negligible because the speed of electromagnetic signals

through a fiber is typically c/n, where c is a speed of light in vacuum and n is a refractive index of a fiber. The functional expression for this propagation delay is $t_p = nx/c$, where x is a distance between nodes. Hence, the actual time for circuit establishment is less than t_a.

2.7   Summary

The routing deflection network was briefly mentioned at the end of the preceding chapter, and it was brought up again at the beginning of this chapter with extensive detail. This technique is probably one of the most practical applications in communication networks to resolve output contention. For instance, it can be used to lighten telephone congested traffic. The reason that VCPS scheme was chosen is that the response time of a bacteriorhodopsin film is not sufficient for packet switching. However, this scheme capitalizes on both circuit and packet switching technologies. It is very advantageous for a long data stream application, i. e., voice/video/data communication, which is the illustration given in this chapter. The proposed system consists of a crossbar switch and routing controller. Functional description of each processor in the controller was provided in a deterministic manner according to its block diagram. The diagram clearly shows a logically operational function, so that it can properly address the SLM of a switching device where the operational function of each component was described in full detail. The concept of the free-space interconnections is based on optical interferometry of two write beams, which is discussed in the next chapter.

Chapter 3

Diffraction by Optical Interferometry

Holography has not only introduced a new dimension into optics, it has also touched almost all of the traditional areas of applied optics. Its principle affects all applications of interference of waves. Hence, it is standing tall in free-space reconfigurable optical crossbars, where holographic gratings of a dynamic photorefractive material are needed to set up and tear down rapidly in real time. In addition, the fringe benefits include the use of a monostage system, freedom from internal blocking, large interconnection density, fiber optic bandwidth capacity and data transparent redistribution. Another advantage of the holographic method is the preservation of both phase and amplitude information. Hence, it is possible to read a phase hologram or an amplitude hologram independently.

This chapter discusses the basic principles of holographic gratings by optical interferometry. Holography is a wavefront reconstruction by an interference method. It is mainly concerned with the interference of two plane waves as shown in Sec. 3.1. The purpose of this section is to analytically derive Bragg's equation, including the study of plane and volume hologram formations. The application of the grating equation is illustrated in Sec. 3.2. The investigations in this section include diffraction in the Bragg regime and the Raman-Nath regime. Finally, holographic thickness classification is characterized by employing Q-parameter as shown in Sec. 3.3.

3.1 Interference of Two Plane Waves

Consider a uniform plane wave of an electric field in a phasor form with a positive propagating direction as shown in Equ. (1-1). Consider the polarization of this field to be perpendicular to the plane of incidence. Let angle θ be an oblique incidence to the normal of the incident plane as shown in Fig. 3-1. Hence, the polarization vectors $\mathbf{n}_\ell$ in terms of incident angles are defined as :

$$\mathbf{n}_1 = \mathbf{a}_x \sin \theta_1 + \mathbf{a}_z \cos \theta_1 \tag{3-1}$$

$$\mathbf{n}_2 = -\mathbf{a}_x \sin \theta_2 + \mathbf{a}_z \cos \theta_2 \tag{3-2}$$
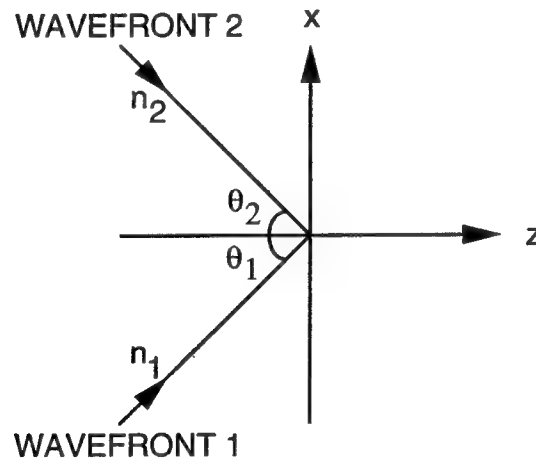
Fig. 3-1 Two-plane waves's interferometry.

The complex amplitude of the electric field can be further expressed in terms of a phase constant ($\beta = 2\pi / \lambda$) as follows :

$$\mathbf{E}(x, z, t) = \mathbf{E}_0 e^{-i(\beta \mathbf{n}_\ell \cdot \mathbf{r} - wt)}, \quad \text{where} \quad \ell = 1, 2 \tag{3-3}$$

In holographic grating applications, it is necessary to consider coherent light in addition to coherent length between the two write beams. Two beams are coherent if and only if they both originate from the same source. These beams would then interfere when they are added together. On the other hand, beams from multiple light sources would not interfere on a long-term basis and their intensities would be the result of a direct addition since they are incoherent. Notice that a beam splitter is one of the necessary tools to divide a single beam from a coherent light source into two coherent plane waves at an angle (θ_1 + θ_2) apart from each other. In this case, both plane wavefronts have the same phase constants in which β_1 equals β_2 for a pure monochromatic plane wave. Since both phase constants are the same, then a single symbol β can be used to represent both constants.

Both fields must be polarized in the y direction to be perpendicular to the plane of incidence which is in the xz plane. The equations representing both incident fields can be written with the substitution of Equ. (3-1) and Equ. (3-2) into Equ. (3-3), which leads to the following equations :

$$E_1(x, z, t) = a_y E_{o1} e^{-i(\beta(x\sin\theta_1 + z\cos\theta_1) - \omega_1 t)} \tag{3-4}$$

$$E_2(x, z, t) = a_y E_{o2} e^{-i(\beta(-x\sin\theta_2 + z\cos\theta_2) - \omega_2 t)} \tag{3-5}$$

Let E(x,z,t) be the resultant complex amplitude which is the sum of both fields. Thus, the functional form in a mathematical representation is $E(x,z,t) = E_1(x,z,t) + E_2(x,z,t)$. The detailed expression of this function is:

$$E(x, z, t) = a_y E_{o1} e^{-i(\beta(x\sin\theta_1 + z\cos\theta_1) - \omega_1 t)} + a_y E_{o2} e^{-i(\beta(-x\sin\theta_2 + z\cos\theta_2) - \omega_2 t)} \quad (3\text{-}6)$$

It has been a tradition to detect an intensity associated with the resultant field rather than just the field itself. Therefore, the resultant intensity becomes:

$$I(x, z, t) = E(x, z, t)E^*(x, z, t) = (E_1(x, z, t) + E_2(x, z, t))(E_1^*(x, z, t) + E_2^*(x, z, t))$$

$$I(x, z, t) = |E_{o1}|^2 + |E_{o2}|^2 + E_{o1}E^*_{o2} e^{-i(\beta x(\sin\theta_1 + \sin\theta_2) + \beta z(\cos\theta_1 - \cos\theta_2) + (\omega_2 - \omega_1)t)}$$
$$+ E^*_{o1}E_{o2} e^{i(\beta x(\sin\theta_1 + \sin\theta_2) + \beta z(\cos\theta_1 - \cos\theta_2) + (\omega_2 - \omega_1)t)} \quad (3-7)$$

where v is a phase velocity of an electric field, the angular frequency ω_1 equals ω_2, such that $\Delta\omega = \omega_2 - \omega_1 = 0$. Since $\beta = \omega/v$, then Equ (3-7) is time independent. Let 's consider z to be a constant: then the second term in the argument of the exponential function is merely a constant-phase contribution. Notice that the relationship between a wavelength in the x-direction and its phase constant is $\lambda_x = 2\pi / \beta_x$, and the phase constant in the x-direction of Equ. (3-7) is $\beta_x = \beta(\sin\theta_1 + \sin\theta_2)$. In general, the relationship between the phase constant to the wavelength of a propagating plane wave is defined to be $\beta = 2\pi / \lambda$. Therefore, the substitution of β's equation to β_x's equation induces λ_x's equation as follows:

$$\lambda_x = \frac{\lambda}{\sin\theta_1 + \sin\theta_2} \quad (3\text{-}8)$$

The wavelength λ_x is the period spacing of the sinusoidal intensity distribution commonly known as the grating spacing Λ ($\Lambda = \lambda_x$). Thus, Equ. (3-8) is known as the grating equation [49, 50] which can be rewritten in a more familiar form as follows:

$$\Lambda(\sin \theta_1 + \sin \theta_2) = \lambda \tag{3-9}$$

If the incident angles of both plane wavefronts are exactly identical to each other, such that angle θ_1 equals θ_2, then Equ. (3-9) becomes Bragg's equation which is completely in agreement with the geometrical result that was given in reference 38. Thus, the expression of Bragg's equation of Equ. (3-9) becomes

$$2\Lambda \sin \theta = \lambda \tag{3-10}$$

where $\lambda = \lambda_a / n$, and λ_a is the wavelength in free space and n is the average of the refraction index of a photosensitive medium.

Let's consider an idealistic experimental condition where both incident angles are equal to an angle θ ($\theta_1 = \theta_2 = \theta$). Thus, it reduces Equ. (3-7) to a one-dimensional spatial function I(x) as follows:

$$I(x) = \left|E_{01}\right|^2 + \left|E_{02}\right|^2 + E_{01}E^*_{02}\, e^{-i2\beta x \sin \theta} + E^*_{01}E_{02}\, e^{i2\beta x \sin \theta} \tag{3-11}$$

Invoking the definition of intensity distribution, with assistance from Euler's famous formula, Equ. (3-11) can be further simplified to a sinusoidal oscillation form. But this can only be achieved after the treatment of the constant multipliers belonging to the third and fourth term in the equation. These multipliers can be expressed in an intensity term as follows:

$$E_{01}E^*_{02} = \sqrt{(E_{01}E^*_{02})(E^*_{01}E_{02})} = \sqrt{\left|E_{01}\right|^2 \left|E_{02}\right|^2} = \sqrt{I_{01}I_{02}} \tag{3-12}$$

Similarly,

57

$$E^*_{01}E_{02} = \sqrt{I_{01}I_{02}} \qquad (3\text{-}13)$$

Hence, the interference law [51] of Equ. (3-11) can be written in a more traditional form as follows:

$$I(x) = I_{01} + I_{02} + 2\sqrt{I_{01}I_{02}} \cos(2\beta x \sin\theta) \qquad (3\text{-}14)$$

where I_{01} and I_{02} are the intensities of plane waves 1 and 2, respectively. The ratio of the resultant intensity $I(x)$ to its time average distribution $(I_{01} + I_{02})$ is called the distribution intensity $I_d(x)$, which can be expressed in terms of grating spacing by invoking the definition of phase constant $(2\beta\sin\theta = 2\pi/\Lambda)$. Hence, the distribution intensity becomes:

$$I_d(x) = \frac{I(x)}{I_{01} + I_{02}} = 1 + 2\frac{\sqrt{I_{01}I_{02}}}{I_{01} + I_{02}} \cos\left(\frac{2\pi}{\Lambda}x\right) \qquad (3\text{-}15)$$

where $2\dfrac{\sqrt{I_{01}I_{02}}}{I_{01} + I_{02}}$ is known as fringe visibility, which is the contrast of the intensity standing wave pattern.[49,52] Furthermore, the visibility is known as an intensity standing wave ratio. If both intensities are equal $(I_{01} = I_{02})$ then the fringe visibility is unity which is an indication of a perfect spatial coherence. Therefore, the fringe visibility is a tool to measure the degree of coherence.

Coherence refers to the degree of correlation of light from a single source over time and space, hence, the names spatial and temporal coherence which are the two properties which are important in a coherent optical system. Spatial coherence is a measure of light interference when there is no time delay to be considered. Temporal coherence is a measure of light interference when there is a time delay between two wave trains that emanated from the same source. Consequently, temporal coherence is strongly affected by

an unequal path length which degrades the quality of the fringe contrast. An important property in holographic interferometry is coherence length which is a physical path difference (ΔL) between two write beams ($\Delta L = z_2 - z_1$), where z_2 and z_1 are the optical lengths of wavefront 1 and 2, respectively. Furthermore, the coherence length of a radiation source is related to the spectral width or line width ($\Delta \lambda$) by $\Delta L = \lambda^2/\Delta \lambda$, where λ is the center wavelength.[53] For two wavefronts to interfere in an interferometric system, the maximum optical path difference ($z_2 - z_1$) must be less than the coherence length of the source radiation ($\lambda^2/\Delta \lambda$). However, if $z_2 - z_1 > \lambda^2/\Delta \lambda$, then there is no interference between the two wavefronts of a single source. In fact, two beams in this case become incoherent, and their intensities add directly. Consider an incoherent single light source. Equation (3-14) can be rewritten as $I(x) = I_{01} + I_{02} + I_{12}$, where I_{12} is a mutual intensity. If $I_{12} = 0$, then $I(x) = I_{01} + I_{02}$ which is a direct addition of their intensities. Hence, the system is incoherent. Therefore, coherence length is a tool to measure how accurate the two optical path lengths in an interferometry must be balanced.

The grating amplitude is not critically sensitive to a small intensity deviation between two interfering plane waves. Numerous computer simulations for various incremental differences between I_{01} and I_{02} were plotted and shown in Fig. 3-2. The use of equation (3-15) makes it feasible to zoom in on the intensities alone, assuming both incident angles to be Bragg matched. In other words, this figure shows the properties of plane diffraction gratings which are known as plane holograms. The optimal intensity distribution occurs when I_{01} equals I_{02}. As the intensity deviation ($I_{02} - I_{01}$) between both beams increases, the grating amplitude degrades to $(100\dfrac{2\sqrt{\Delta}}{\Delta + 1})\%$ from the maximum fringe visibility, where Δ is the ratio of two write intensities ($\Delta = I_{01}/I_{02}$). For instance, if the intensity deviation is 100%, which I_{02} equals $2I_{01}$, then the grating amplitude reaches

94.28% of its maximum value. Hence, if both intensities of the interfering plane waves were differed by a small order of magnitude then the grating degradation affect is barely noticeable.
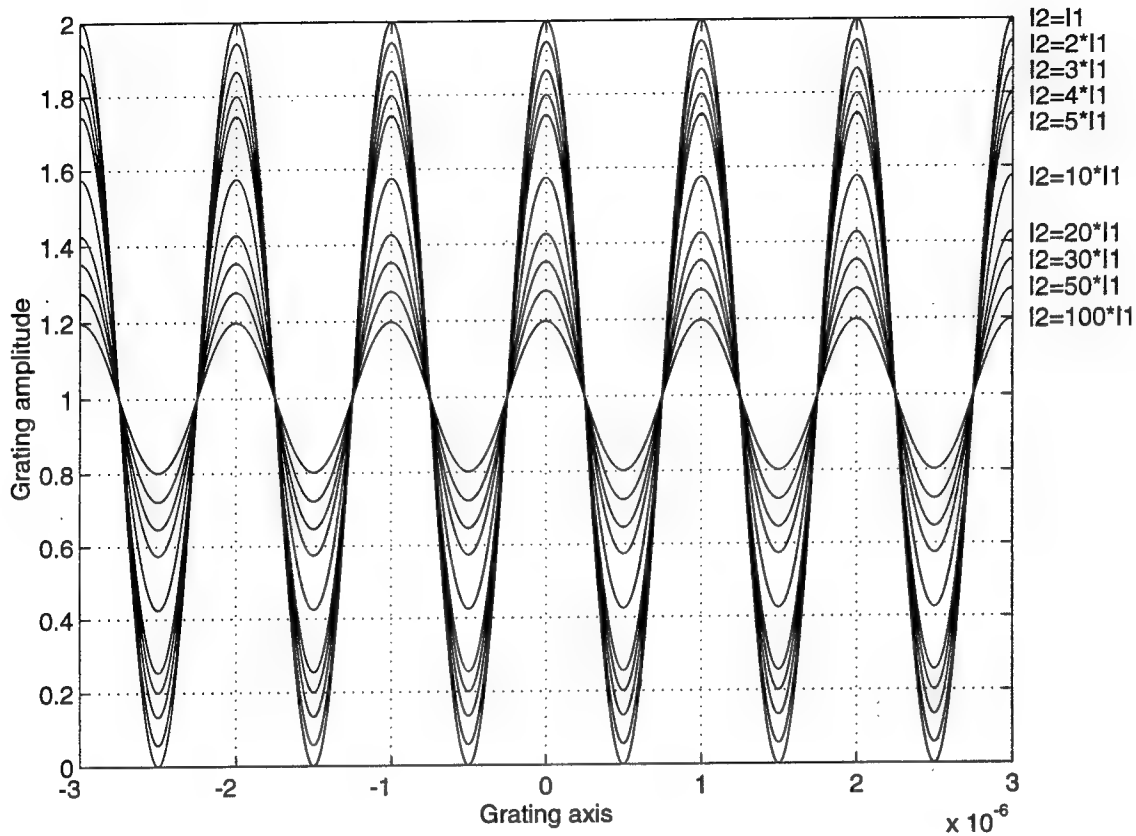
Fig. 3-2 Amplitude grating profile for various plane holograms.

Let's consider Equ. (3-7) under two practical conditions where an incident angle θ_1 is slightly different from angle θ_2, and both write intensities have different values. These conditions are set forth to emulate a real experimental situation which will provide a sense of understanding about a realistic circumstance where it is extremely difficult to have a precise and accurate duplicated image between two writing beams. Thus, the sinusoidal oscillation intensity in this case becomes:

$$I(x,z) = I_{01} + I_{02} + 2\sqrt{I_{01}I_{02}} \cos[\beta x(\sin\theta_1 + \sin\theta_2) + \beta z(\cos\theta_1 - \cos\theta_2)] \qquad (3\text{-}16)$$

Notice that the spacing grating term in the cosine argument can be replaced by $2\pi/\Lambda$ which is the result of a 2π multiplication to both sides of Equ. (3-9) and rearrangement of the resultant equation yields $\beta(\sin\theta_1 + \sin\theta_2) = 2\pi/\Lambda$.

Let δ be an angle representing a different value between θ_1 and θ_2. Furthermore, consider this angle to be very small qualifying for a small angle approximation. Consider angle θ_2 to be the sum of θ_1 and δ ($\theta_2 = \theta_1 + \delta$). After replacing angle θ_2 by $\theta_1 + \delta$ in Equ. (3-16), and invoking trigonometric identities on addition formulas of sine and cosine in addition to power series expansions on $\sin\delta = \delta$ and $\cos\delta = 1$, the resultant equation can be algebraically simplified to a functional form as follows:

$$I_d(x, z) = 1 + 2\frac{\sqrt{I_{01}I_{02}}}{I_{01} + I_{02}} \cos\left(\frac{2\pi}{\Lambda}x + \frac{2\pi n}{\lambda_a}\delta(x\cos\theta_1 + z\sin\theta_1)\right) \qquad (3\text{-}17)$$

where θ_1 is Bragg's angle. There are two physical interpretations of angle δ. First, if this angle equals zero ($\delta = 0$) then there is no grating depth and the interference fringes are plane diffraction grating intensity distribution in the x-plane. Hence, Equ. (3-17) becomes Equ. (3-15) which is the equation of perfect phase coherence. Second, if angle δ is not equal to zero ($\delta \neq 0$) then the grating profile is no longer just at the surface but also has depth along the z-axis. This is a property of a volume diffraction grating which is commonly known as a volume hologram. Furthermore, the significance of δ's value is an indication of how far Equ. (3-17) departs from a Bragg condition.

Let's remove restrictions on a small angle δ that bounds it to a small angle approximation. Hence, Equ. (3-16) can only be simplified by brute force to this functional form:

$$I_d(x,z) = 1 + 2\frac{\sqrt{I_{01}I_{02}}}{I_{01}+I_{02}} \cos[\frac{2\pi}{\Lambda}x + \frac{2n\pi}{\lambda_a}z(\cos\theta_1 - \cos\theta_2)] \qquad (3\text{-}18)$$

Notice that $\theta_2 = \theta_1 + \delta$. A significant change in this equation is its two-dimensional spatial function, whereas Equ. (3-15) represents a photosensitivity in only one dimension. The function of Equ. (3-18) does not only describe an areal grating intensity profile but also a grating depth which is governed by z-term characteristics in the cosine's argument. The values of z vary from zero to a maximum thickness of a sample. Both equations (3-15) and (3-18) are identical at the surface of a sample where z is equal to zero. But Equ. (3-18) can be used to describe the intensity profile of any holographic materials that have photosensitive volume. This equation produces volume diffraction gratings or volume holograms.

The effects of a small δ-angle have a great influence on the orientation of the fringes with respect to the film normal. Computer simulations of Equ. (3-18) illustrate the importance of these effects. As an example, consider the following parameters: A sample has a 100-μm thickness and an average index of refraction of 1.551. The periodic spacing grating in the sample is 1μm which is written by two equal intensity beams at a 413-nm wavelength in free-space. An interbeam angle for various δ-angles can be determined by utilizing Equ. (3-9). The crest fringes of Equ. (3-18) vary cosinusoidally as functions of x and z. All four volume grating profiles were generated by computer simulations. Both two
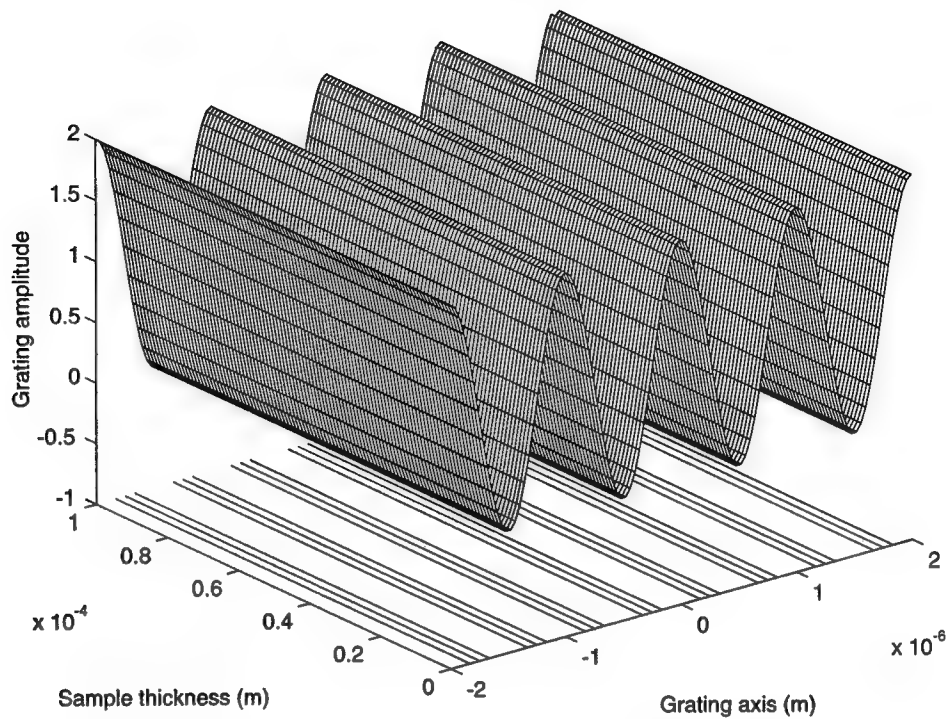
Fig. 3-3 Volume grating profile for 0° δ-angle.
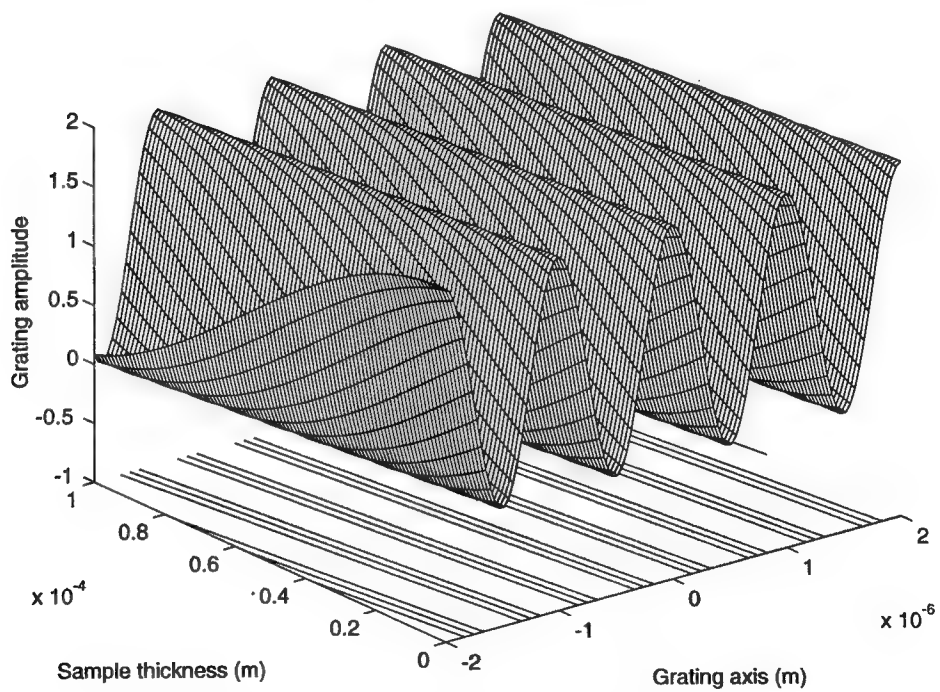
Fig. 3-4 Volume grating profile for 0.5° δ-angle.
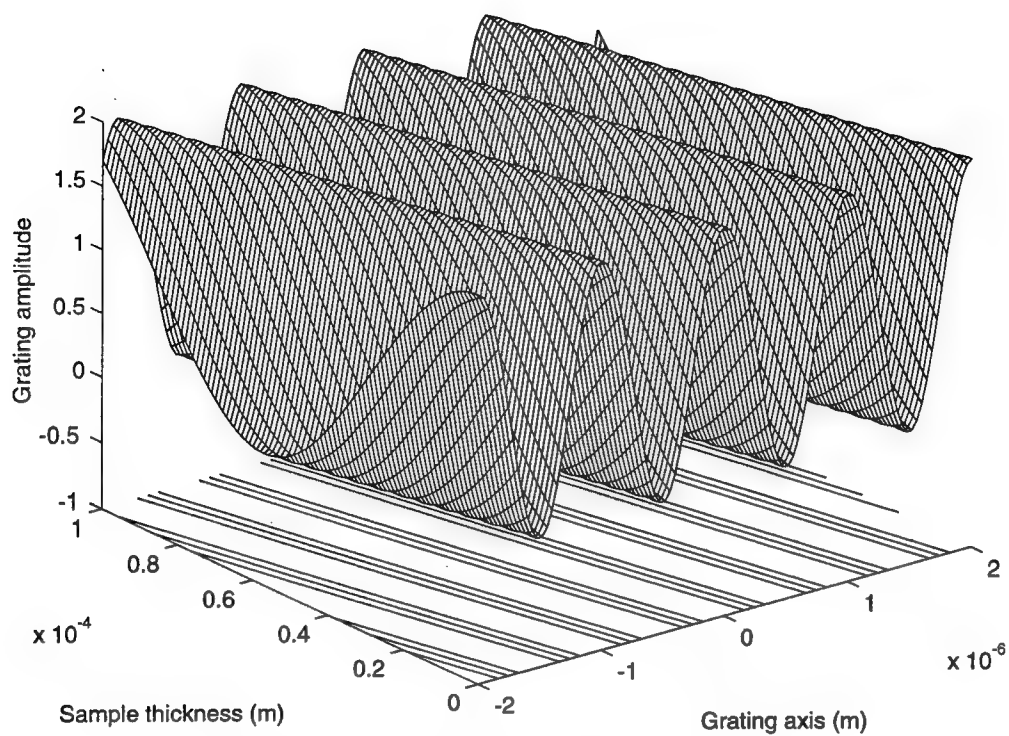
Fig. 3-5 Volume grating profile for 1° δ-angle.
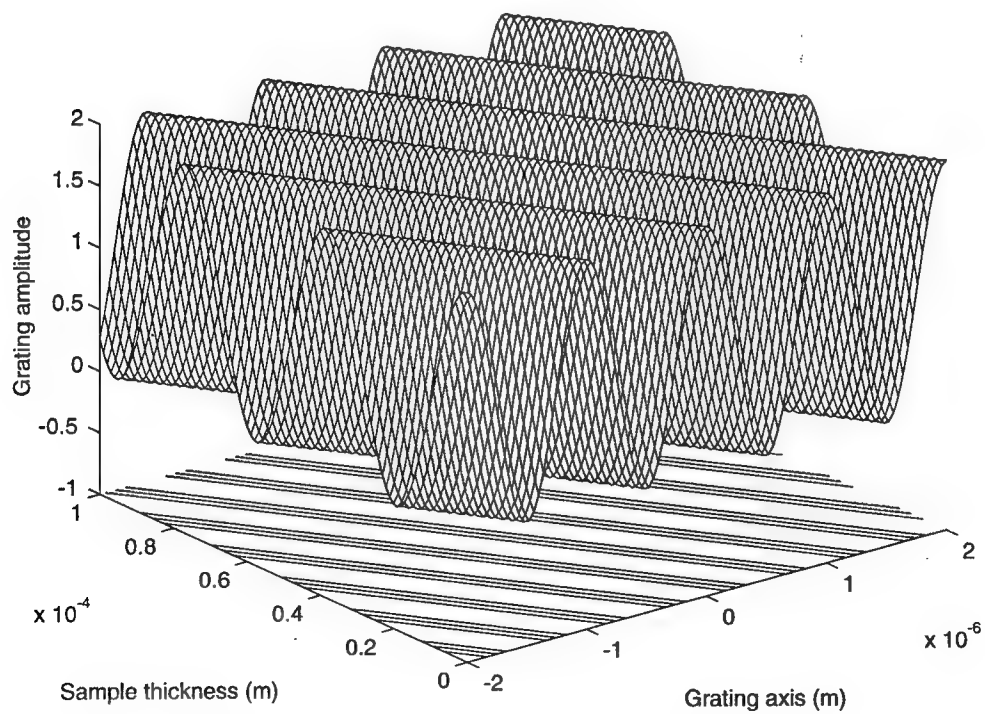
Fig. 3-6 Volume grating profile for 3° δ-angle.

and three dimensional grating profiles are shown in all four figures. Their purpose is to demonstrate that orientation of gratings is changed with respect to the film normal according to δ-angle. A value of δ-angle approaches zero if and only if both incident angles have the same value. In this case, volume diffraction gratings are formed parallel to the surface normal. If δ-angle has an absolute value exceeding zero then the grating orientation becomes slanted with respect to the surface normal. A value of δ-angle can be controlled either by incident angles or by rotation of a sample or by both methods to minimize angle deviation. These figures have illustrated that the degree of slanted fringes depends on how large the δ-angle is, as shown in Fig. 3-3 through Fig. 3-6 which have δ-angles $0°$, $0.5°$, $1°$ and $3°$, respectively.

3.2   Application of Grating Equation

There are a number of interconnection systems that are useful in optical routing, but the foreseen journey of this research will only embark on a free-space reconfigurable holographic crossbar where routing switches are controlled by volume diffraction gratings. These gratings are created by the interference of two monochromatic plane waves. This method can produce both transmission and reflection holograms which are the main focused interest in this research. The laser is probably the only light source in nature to emit coherent light, but no two lasers are identical. Therefore, two write beams must emanate from a single light source, and the maximum optical path difference must not exceed $\lambda^2/\Delta\lambda$ in order to comply with a coherence length. The constructive interference between these two write beams creates volume diffraction gratings with periodic spacing Λ in a holographic medium. Normally, an incident angle can be easily calculated by

assuming that an incident interbeam angle between two write beams is 2θ, where θ is the Bragg angle which is an angle formed by a write beam and the surface normal of a sample. In fact, the Bragg angle is usually determined for a known grating period Λ. The tool to determine this angle is called Bragg's law which is given by Equ. (3-10). Solving this equation yields a Bragg angle to be:

$$\theta = \arcsin(\frac{\lambda_{w,a}}{2n_w \Lambda})\qquad\qquad(3.19)$$

where $\lambda_{w,a}$ is the wavelength of a write beam in free space, and n_w is the average refractive index of a holographic medium corresponding to wavelength $\lambda_{w,a}$. A fundamental trigonometric identity like the tangent function is needed to configure a geometry of write beams after the Bragg angle has been computed. At this junction, all information has been determined for an experimental setup to write a specific grating.

After the diffraction gratings have been written for a specific incoming-to-outgoing mapping, modulated data over a known carrier (read beam) will be diffracted to a desired output port. The term diffraction is referred to both transmission and reflection of a beam in which it bends from a rectilinear path after it has struck on gratings. The incidence of a read beam must enter at its Bragg angle in order to capitalize on its maximum diffraction efficiency. If the read and write beams have two distinct wavelengths, then the incident angles are also different. Hence, Equ. (3-19) must be modified to incorporate both wavelength and refractive index of a read beam before computing its Bragg angle.

In general, the diffraction angle of a read beam is not necessarily equal to its incident angle. The equation for a first order diffraction was given by Equ. (3-9). This grating equation can be modified to accommodate all high diffraction orders by a phase grating that splits the incident beam into various orders. Thus, Equ. (3-9) becomes:

$$\sin \theta_m = m \frac{\lambda_{r,a}}{n_r \Lambda} - \sin \alpha_i \qquad (3\text{-}20)$$

where the integer m is a diffraction order number (i.e., $m = 0, \pm 1, \pm 2, \pm 3, ...$). The symbol $\lambda_{r,a}$ is an optical carrier wavelength, n_r is an average refractive index of a holographic medium at that wavelength $\lambda_{r,a}$, α_i is an incident angle of the read beam, and θ_m is the angle of the mth-order diffraction beam. Notice that the angle separation between two adjacent neighboring orders is $\lambda_{r,a}/n_r \Lambda$ in a sample.

The expression of Equ.(3-20) somewhat represents a general condition for an oblique incidence in the first quadrant and an obtuse incidence in the second quadrant because sine is a positive function in both quadrants. Follow general convention, an incident angle in the second quadrant as shown in Fig. 3-7 should be a supplementary angle. This angle is a supplement of an obtuse incidence ($\angle$ 180 - α_i). Thus, this equation possesses the proper sign convention for both incidences, first and second quadrant, because the sine function yields sin (α_i) = sin (180 - α_i) for an angle α_i greater than or equal to 0° but less than or equal to 90°.

Consider the lower and upper bound of a sine function which has its value between -1 and +1, respectively. This self-imposed restriction confines the right hand side of Equ. (3-20) between these two boundaries as follows:

$$-1 \le m \frac{\lambda_{r,a}}{n_r \Lambda} - \sin \alpha_i \le 1 \qquad (3\text{-}21)$$

Applying a simple algebraic manipulation to this equation yields the functional form that discloses a range of a diffraction order for each given spacing grating Λ. Thus, this range is governed by:

$$\frac{n_r \Lambda}{\lambda_{r,a}}\left(\sin \alpha_i - 1\right) \le m \le \frac{n_r \Lambda}{\lambda_{r,a}}\left(\sin \alpha_i + 1\right) \tag{3-22}$$

This expression reveals that there is always a zeroth order diffraction in interferometry because zero is always one of the numbers in the set that describes a range. It is obvious that the direction of the zeroth order diffraction coincides to its incident path for a transmission hologram. The zeroth order diffraction of a reflection hologram occurs at an angle that has the same value as an incident angle ($\theta_0 = \alpha_i$). A diffraction set is a set of integers that always include zero. The set contains either positive or negative integers or both depending on how the phase grating has split an incident beam into various orders as shown in Fig. 3-7 which is a multiple diffraction order.[54] This figure represents an optical
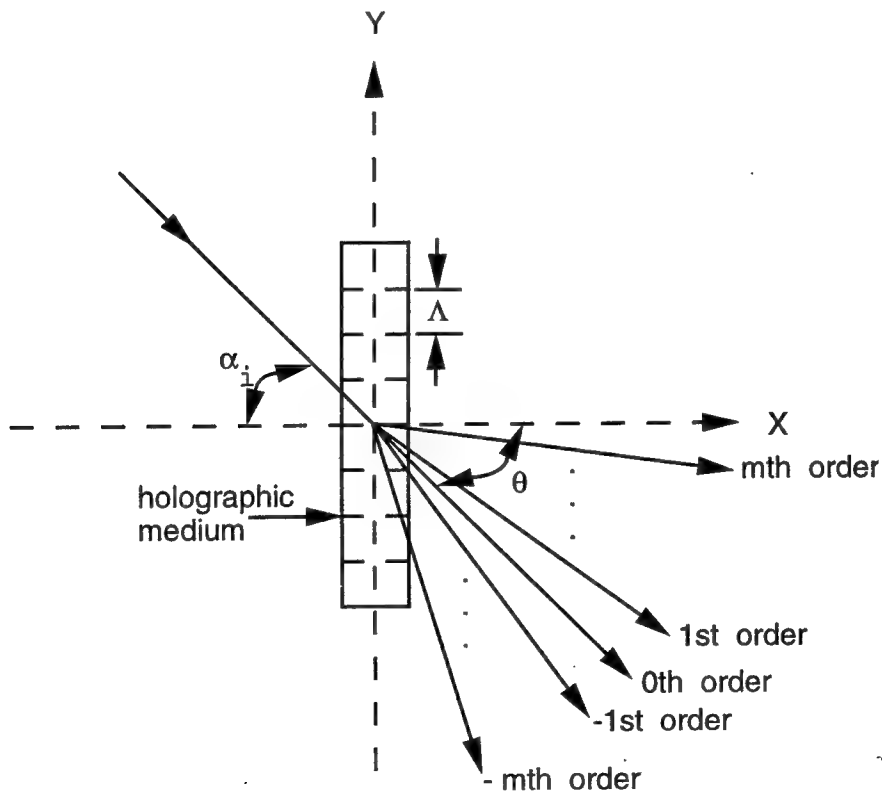
Fig. 3-7 Multiple diffraction orders in the Raman-Nath regime.

interferometric system that is operating in the Raman-Nath diffraction region. This region is characterized by the simultaneous generation of a multiple diffraction order. In this case m is greater $\pm$ 1. Note that an experimental result has shown that a low positive order diffraction is always brighter (higher diffraction efficiency) than a higher order one. Furthermore, if a grating Λ is sufficiently small then the system is always operating in the Bragg regime, in which no diffraction order is higher than the first order diffraction. In addition, the maximum diffraction efficiency can be achieved if and only if an incident beam is along the Bragg angle. However, off-Bragg diffraction is possible but their efficiencies are greatly compromised.

3.3 Holographic Thickness Classification

How does the thickness of a hologram impact on which theory to apply ? This question can be answered by first introducing different names (thin hologram and thick hologram) for these holograms that may unlock an explanatory door. Another name for a thin hologram is a plane hologram, whereas the thick hologram is known as a volume hologram. These two holograms are usually represented by two different sets of equations. For instance, Equ. (3-15) can only be used to describe the grating profile of a plane hologram, whereas Equ. (3-18) reveals the description about diffraction gratings of a volume hologram. Therefore, a criterion must be established to classify a hologram for plane or volume theory to be applied. A large amount of work has been done on acoustic diffraction gratings, which is in many aspects similar and can directly apply to hologram gratings. The parameter Q was one of the products of this work.[55, 56, 57, 58] The holographic application of this parameter is used to classify the type of a hologram. The parameter Q has been defined as follows:

$$Q = \frac{2\pi\lambda_a d}{n\Lambda^2} \tag{3-23}$$

where d is a thickness of a sample. A hologram can be regarded as thin if it satisfies the condition of Q < 1. For a thick hologram, the parameter Q is much greater than one (Q > 1). Thus, when Q ≈ 1, it gives a transition region between the two holograms. One can conduct further investigation to determine the absolute minimum Q's value for a volume hologram. One study has shown that the coupled wave theory for thick hologram gratings begins to produce good results at Q ≥ 10.[59] Let's consider some typical values of a bacteriorhodopsin film. Suppose these values are: d = 100 μm, λ_a = 570 nm, n = 1.537. If Q equals 10 (Q = 10), substituting these values into Equ. (3-23) produces a 4.827-μm fringe period (Λ=4.827 μm). Applying this periodic value into Equ. (3-19) yields the value of the Bragg angle to be 2.2°, then the minimum interbeam angle needed to create the volume hologram is 4.4°.

Two figures are introduced below to further illustrate the relationships of the parameter Q to the Bragg angle, as well as the fringe spacing period and thickness of a sample. The relation of Bragg angle versus parameter Q for various thicknesses, as shown in Fig. 3-8, illustrates that thickness of a sample imposes a restriction on the upper limit of parameter Q. For instance, to keep the Bragg angle less than 90° the parameter Q must not exceed 67 for a sample of 1-μm thickness and written by laser beams that have 570-nm wavelength in free-space. It should also be pointed out that a large Q produces a large Bragg angle, but this angle becomes smaller as thickness of a hologram increases. Gratings can be written in a sample after the Bragg incidence has been determined. Figure 3-9 shows the relationship between the spacing grating period and the thickness of a

sample for various Q-parameters. It reveals that increasing the thickness of a sample will enhance the fringe spacing period for each Q's value corresponding to a specific Bragg angle.
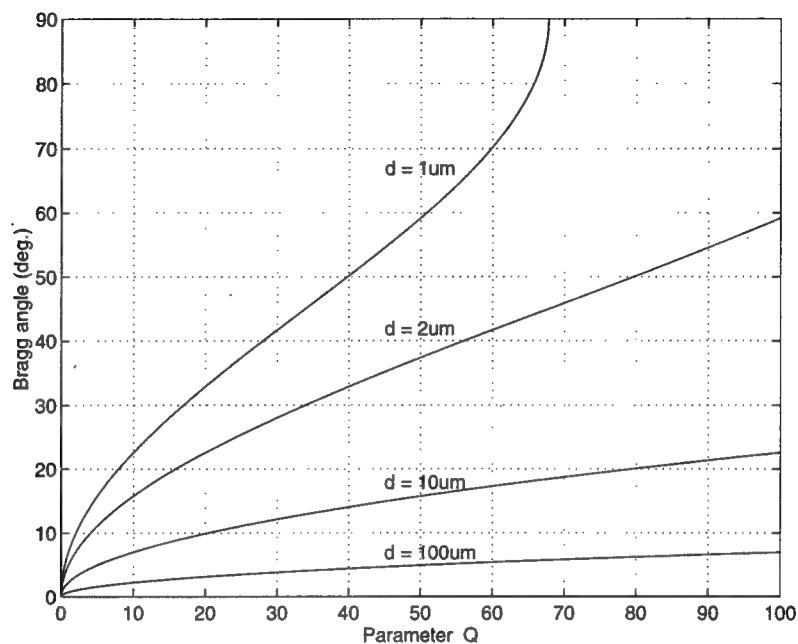
Fig. 3-8 Bragg angle vs. Q-parameter for various BR thicknesses.
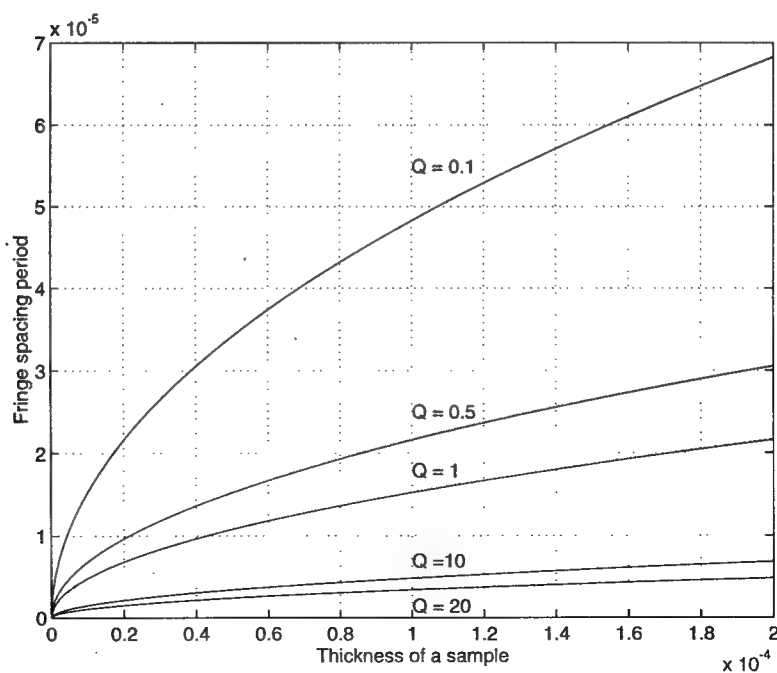
Fig. 3-9 Fringe spacing period vs. thickness of a sample for various Q-values.

71

3.4 Summary

The approach of this chapter was to derive Bragg equation by two-plane wave interferometry. Bragg angle is a fundamental parameter of a hologram study. Hence, it is necessary to examine various holographic grating profiles as a function of write intensity variations and incident angle differences. The grating equation was utilized to determine the Bragg diffraction region and the Raman-Nath diffraction regime. This chapter ends with the characterization of a photorefractive polymer thickness by the Q-parameter. The polymer used in this switch design is a bacteriorhodopsin based system. The description and properties of this material are explored in detail in the proceeding chapter.

Chapter 4

Complex Photochemical Reactions in a Bacteriorhodopsin Film

Bacteriorhodopsin (BR) is a light transducing protein in the purple membrane of a salt-loving microorganism. It has a strong absorption in a broad region of the visible spectrum. The B-state in the photocycle can be considered to be the ground state which has absorption maxima at 570 nm, whereas the intermediate M-state has the longest thermal lifetime with its absorption maxima at 413 nm. Moreover, BR can be genetically or chemically modified to meet different application requirements. For instance, it can be mutated to a D96N variant (BR_{D96N}) by replacing aspartic acid D96 with asparagine to prevent the rapid intramolecular reprotonation of the Schiff base. Perhaps, the most intriguing features of this organic photopolymer are its extraordinary stability in the chemical, thermal and photochemical sense, its large optical nonlinearity, its dynamic nature, and its durability. Furthermore, BR-doped polymer film can be fabricated for a large scale crossbar application, whereas photorefractive crystals like BSO or KNSBN cannot be grown easily to the same dimension as BR. In addition, BR is a thin film which in phase holograms removes the Bragg angle restriction.

Complex photochemical reactions in the film are discussed in this chapter. Description of salt-loving microorganisms is given in Sec. 4.1. Sec. 4.2 provides an explanation of a reversible reaction scheme. Transition rates by one and two-photon processes are analytically studied in sections 4.3 and 4.4, respectively. Sec. 4.5 is devoted to the theoretical investigation of transient dynamic holograms.

4.1 Salt-Loving Microorganisms

Life can be found in the most unlikely places where the environment is naturally inhospitable such as in low oxygen content, high concentration salt marshes. When intense sunlight, exceeding twice the temperature of the human body ($> 74°C$), illuminates shallow seawater for an extended period of time causing oxygen to evaporate, it induces seawater to salt marshes that have salt content roughly seven times that of ordinary seawater. In other words, these marshes have a high density of sodium chloride, low oxygen concentrations, and high light intensity. This harsh environment becomes a suitable habitat for an unusual colony of bacteria. A great concentration of this bacterial population changes a salt marsh coloration to deep red which is one of the membrane colors. Membrane is the door into a cell. The red color is due to one of the pigments in the cell.[60] These rod-shaped bacterium have dimensions about five to ten microns in length and approximately 0.7 microns in diameter. The bacterium is a protein-based microorganism called Halobacterium salinaritum which is commonly known as Halobacterium halobium.[61] The photosynthetic mechanism that converts sunlight to chemical energy in this specie was first discovered by Walther Stoeckenius in 1964.[62] Stoeckenius et. al. have successfully fragmented (centrifugation) these rod-shaped cells into three different colors. They have learned that the orange-red group belongs to the low buoyant density. Its size and composition are random. The function of the red pigment is to safeguard against the blue light high intensity affect. Size and composition of other two groups have a high degree of uniformity. Yellow pigment can be found almost entirely on the walls of gas-filled sacs in the protoplasm of the cell. This group has the highest buoyant density by comparison to the other two groups. Hence, its function seems to maintain a certain floatation depth of the cell. The last group belongs to the most important membrane. This membrane has purple pigment, and its density of buoyancy is high due to about 25 percent of lipid content and the remainder is protein of a single specie. The purple membrane is an independent

74

membrane cell in a lipid matrix that has protein to lipid ratio of 3:1. The function of this purple membrane seems to convert light into an energy source to power metabolic processes without a respiration operation in the oxygen deficient environment. A substantial proportion of hydrophobic (water-avoiding) amino acids was found in the protein portion. The purple membrane seems to have a molecular weight of about 26000 or one mole of retinal per 26000 grams of protein. The cells have a large concentration of purple membrane which covers up to 50 percent of the surface of a membrane cell. Further inspection has revealed that a purple membrane has an ovally round sheet-like configuration with a dimension about a half micron in diameter, and structure of both surfaces appears to be different.[63] However, the arrangement of the protein base is a regular two dimensional hexagonal crystalline lattice in the plane of the membrane with lipid filling in the gaps.[63] In addition, the structure and composition of the chromoprotein in the membrane is very similar to the visual pigment in animal eyes. This intriguing similarity may have prompted Oesterhelt, who joined Stoeckenius in 1969, to investigate a correlation between the purple membrane protein and a rhodopsinlike chromoprotein. Rhodopsin is also known as visual purple, a biological class of a special protein (opsin) with a red light sensitive pigment in the retinal rods of animal eyes. Retinal possesses a broad absorption band around 380 μm, ergo it is a yellowish to orange aldehyde of vitamin A which forms a visual pigment if it were to combine with a protein. He found that the purple membrane protein contains one retinal group per molecule. Hence, they named this light transducing protein in the purple membrane **bacteriorhodopsin** which is a critical membrane for these microorganisms to survive in the prescribed environment. In short, the function of bacteriorhodopsin (BR) is to absorb light like sensory pigments without triggering a sensory response. Instead it converts that harvested light energy into chemical energy for metabolic processes to sustain life via transmembrane proton pumping by light driven protons.[64] Note that the BR molecule is a retinal protein molecule in a purple membrane cell. Thus, an elucidative description of a light absorbing chromophore (molecule) was given by Birge.[61] A positive

75

charge from a protein-bound chromophore can interact electrostatically with charged amino acids to produce photochemical properties. A molecule's shape is effected by the absorption of light. The light causes positive charges to position themselves perpendicularly to the protein membrane sheet which in turn generates a photoelectric signal by this charge movement, and the cell metabolism is sustained by the energy distribution from the proton transportation across all of the cell membranes. Perhaps one of the most remarkable properties of a BR protein is its extraordinary stability against chemical, thermal and photochemical degradation.

The simplest definition of state or intermediate state used in this text refers to the wavelength at a spectral absorption maxima. A common single letter abbreviation is used to represent this spectral distinction. In the absence of illumination, all protein molecules are sedate in a D state which has an absorption maxima at 548 nm. This state is known as the dark adapted state where both proteins, all-trans retinal and 13-cis retinal, are spontaneously formed.[60, 65, 66] Under yellow illumination these molecules transform from the D state into a photoinduced transition cycle. Note that initial photon absorption only transforms BR molecules from dark adapted state to light adapted state which is a biological active state. Thereafter, a photocycle is formed by light driven protons where they were pumped across all the cell membranes outwardly from the inside (cytoplasmic) to the outside (extracellular) of the membranes through a proton channel which has several angstroms in length.[65] Under steady illumination, protons in the cytoplasmic membrane are depleted and these protons repopulate at the extracellular membrane. This process induces a proton concentration gradient difference that leads to an electric potential generation between the two membranes. In other words, this deprotonation and reprotonation process during a phototransformation produces an electrochemical gradient from light energy.

The term photocycle in this text always refers to a cycle of BR molecules in light adapted state. The main photocycle occurs under a one-photon process where BR

76

molecules undergo a complex transformation from B state, which has an absorption maxima at 570 nm, to intermediate states J, K, L, M_1, M_2, N, O, and back to B state as shown in Fig. 4-1. This figure is a product of photocycle and thermal reactions in a BR film.[67, 68, 69, 70, 71] The forward transformation can only be initiated by photoexcitation. Yellow beam illumination is preferred since this color corresponds to a 570-nm wavelength which is the maximum spectral absorption of the B-state. Using other wavelengths within this spectral absorption band is also possible. A forward reaction by light illumination creates a primary photochemical product at the intermediate J-state, then these excited molecules thermally relax via the rest of the intermediate states back to the B-state where the cycle is completed. The photoreaction from B to J is very fast, i.e., within 500 fsec. Hence, the intermediate J-state may not be a very stable state. The protein in the B-state is all-trans. Illumination by a wavelength around 570 nm phototransforms molecules in a BR film from B to J state, which involves an all-trans to 13-cis photoisomerization of the protonated Schiff base chromophore about the $C_{13} = C_{14}$ bond of a retinal. Protons then translocate across the membrane by thermal reaction (thermal decay), and these molecules eventually relax to the initial B-state after a 13-cis to all-trans thermal reisomerization of the $C_{13} = C_{14}$ bond.[72] Hence, the proteins of the intermediates J, K, L, M_1, M_2 and N are all 13-cis, but the thermal reisomerization from 13-cis to all-trans occurs in O-intermediate. Therefore, both O and B state contain all-trans protein. Note that a molecule relaxing into B state can simultaneously react into a new photocycle.

Proton depopulation from the Schiff base via an aspartic acid D85 into the extracellular channel occurs during the thermal relaxation from L to M_1 intermediate. This process causes a blue shift of the chromophore absorption by 160 nm with respect to that of B state. An aspartic acid D85 near the chromophore binding site accepts protons from the Schiff base before releasing them into the extracellular channel. On the other hand, a thermal transformation from M_2 to N intermediate is possible by aspartic acid D96 which causes reprotonation through the cytoplasmic channel to the Schiff base. An aspartic acid

77

D96 is an internal proton donor located near the cytoplasmic region. Notice that the Schiff base locates half way between the cytoplasmic half channel and the extracellular half channel, and it is a condensation product between an amino (NH_2) group and the aldehyde (CHO) group of retinal.[62] This base can both protonate and deprotonate during a photoreaction.
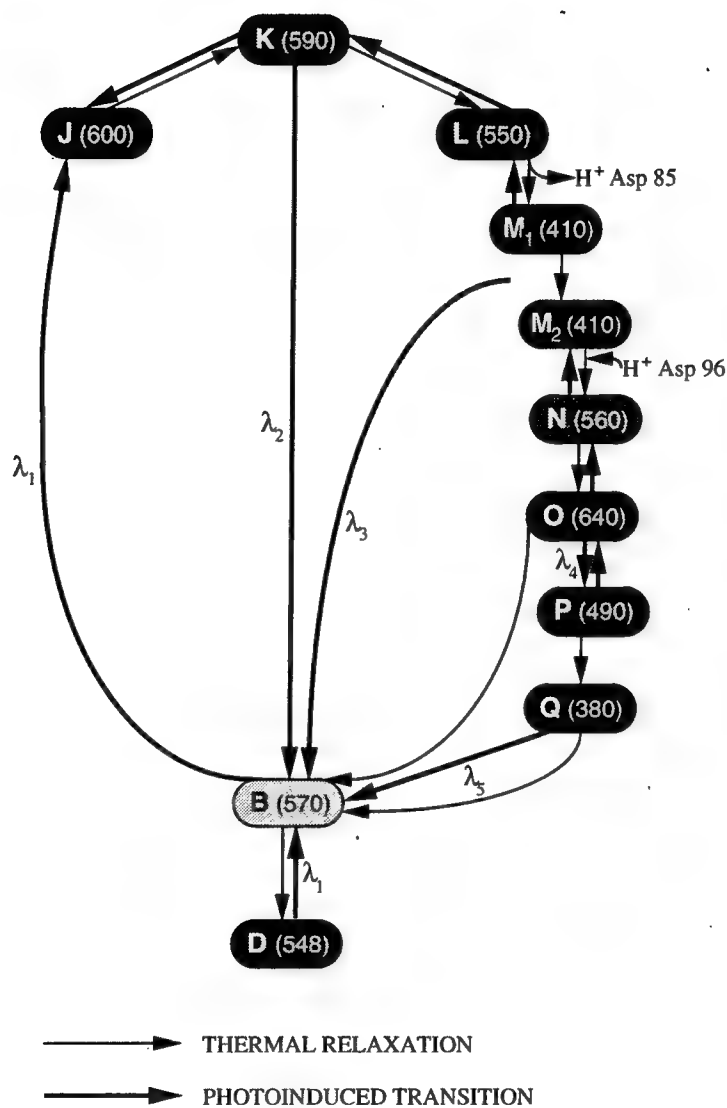
Fig. 4-1 Photocycle and thermal reaction of a bacteriorhodopsin.

The intermediates M_1 and M_2 occupy the same absorption maxima at 410 nm. They both connect in series by an irreversible thermal relaxation with a distinguishable thermal lifetime or kinetic separation.[73] Thus, a single letter abbreviation M is usually used to represent both, M_1 and M_2, thermal states. The intermediate M has the longest thermal lifetime. For instance, the thermal lifetime of K and L intermediate is much shorter than the thermal lifetime of M_1, and the thermal decay lifetime of M_2 is much longer than the rise time of N intermediate. The figure has also shown two disjointed reversed pathways by a photoinduced reversible reaction from intermediate O to N and back to M_2, and also from intermediate M_1 to L to K and back to J which is the last reversible state in the photocycle.

The figure also illustrates two independent phototransitions by a direct phototransformation. These transitions are induced by a two-photon process. The first photon excitation causes a forward transformation from all-trans to 13-cis photoisomerization, and second photon illumination influences excited BR molecules to retransform from 13-cis to all-trans. The terms cis (on the same side) and trans (across) are in reference to, in a simplest definition, a geometrical isomer in which two ligands are found on the same side and opposite side in a molecule, respectively. This two-photon photocycle is also thermally and photochemically stable. Two photocycles by a sequential one-photon process consists of the following states: B → J → K → B,[74, 75] which is induced by yellow and red beam illumination on a BR film, and B → J → K → L → M → B,[76, 77, 78] which is the result of yellow and blue beam illumination on the film.

A forward reaction of BR molecules can be excited by phototransformation with yellow light illumination on a BR film. If a red beam is introduced into a sample colinearly with a yellow light, then red illumination retransforms all molecules in the excited K intermediate back to B state. The result becomes a B-K two-photon photocycle. Likewise, the excited molecules in M intermediate undergo a direct population inversion if they are influenced by blue beam illumination in addition to a yellow light, then a complete pathway

shunts directly from intermediate M back to B state, which in turn produces a B-M two-photon photocycle.

The intermediate O in the main photocycle, which was described in detail previously, can be branched out to form a new cycle called a branching photocycle. This branching photocycle contains the intermediates P and Q.[68, 79, 80] To iterate, species O is formed after thermal reisomerization from 13-cis to all-trans. These all-trans chromophores in O state can again be photoisomerized by a second photon illumination with deep red light. Consequently, this process transforms all molecules in the excited O-state to a 9-cis retinal configuration. They are chromophores of P and Q photoproduct. Albeit, a 9-cis isomer is a product of two sequential one-photon steps which can be viewed as two photon absorption of a linear process.[81] But Birge has referred to it as a sequential one-photon process, because these two photons are not simultaneously absorbed as in a single and nonlinear process. A combination of yellow and deep red light illumination produces a P state, and the molecules in this state gradually relax to the Q state by thermal reaction. The transition from P to Q is only forwardly possible by thermal decay transition of several minutes. The most remarkable property about this branching photocycle is the longevity of the Q state. It has a stable lifetime at room temperature extending to many years. Birge has estimated its lifetime to be about five years. The stable longevity of the Q state is caused by 9-cis chromophores trapped inside a binding site because the bond that links this chromophore to the protein of B-state is broken. However, illumination of deep blue light around 380 nm will quickly isomerize 9-cis to all-trans that will not only induce a spontaneous reconstruction of the broken bond in a chromophore-protein link but also reform the B-state.

The photo and thermal reactions, as shown in figure 4-1, are the photocycle properties of a wild type bacteriorhodopsin film. The wild type refers to a native BR where its photophysical properties have not been altered either by chemical treatment or genetic reengineering. However, this organic material can be mutated to different variants to

optimize its optical properties according to the demand of a specific application. For instance, BR can be mutated to a D96N variant if a long M-state's lifetime is required for an application . This D96N mutant BR, BR_{D96N}, can be obtained by replacing aspartic acid D96 with asparagine in BR, which in turn prevents a rapid intramolecular reprotonation of the Schiff base. Removing aspartic acid D96, an internal proton donor, forces BR_{D96N} to depend on an extramolecular proton availability.[82] Hence, external proton concentrations by pH has a strong influence on the M-lifetime of BR_{D96N}. A high pH value prolongs the M-lifetime because it has low external proton donor concentrations. The relationship between pH and proton concentrations is given by $pH = -\log[H^+]$ which has an inverse proportionality between a parameter pH and $[H^+]$, where $[H^+]$ is a hydrogen ion concentration (proton concentration), and its unit of measurement is the molar (mole per liter).[83] A direct benefit of long M-lifetime is an improvement of diffraction efficiency because the excited molecules tend to be trapped in the M-state causing it to gain population.

Another fascinating bacteriorhodopsin-variant is the blue membrane BR which has a stable lifetime of photoinduced refractive index modulation for several months.[84] This blue membrane BR variant offers an undeniable potential for dynamic holographic data storage applications. The transformation to blue-membrane BR causes it to substantially modify its optical properties into a new species. It no longer possesses the same photo and thermal reaction as the wild type. Instead, it becomes a photochromic holographic material with an optical absorption maxima at 600 nm. In fact, its photocycle is distinctly different from native BR. Furthermore, it can be transformed to a pink membrane by red beam illumination. This new species, pink membrane, has a spectral absorption maxima at 490 nm, and exhibits a very long lifetime. This second species is a photoreversible formation. Hence, the long shelf lifetime of the pink membrane can be cut short by photoreactions. The photoreaction transforms the pink membrane to a blue membrane.

A blue-membrane BR can be produced either by a chemical treatment or a genetic modification. The chemical treatment is a process to either remove or displace divalent cations from a purple membrane which contains about four or five divalent cations. Note that one BR molecule has one calcium and three or four magnesium ions. This treatment will modify BR to a new organic material called deionized or acid blue membrane.[85] The significant contribution of divalent cations is responsibility for maintaining purple color and photochemistry in a BR film. Removal of divalent cations will annihilate the deprotonated Schiff base that creates the M-state. Hence, the affect of this removal does not only alter the photocycle but also shift the spectral absorption maxima.

A genetic mutation of BR is achieved by replacing aspartic acid D85 with asparagine (BR_{D85N}) which completely suppresses a BR's abilities to pump protons across cell membranes, and deprives it of the charged carboxyl group, which in turn, leads to a blue chromophore. Therefore, this BR_{D85N} mutagenesis is called a blue membrane BR. Thus, this type of mutation is permanent, BR_{D85N} can never return to purple at any pH's value.[86] Even so, an activity of proton exchange occurs on the same side of the membrane in a blue mutant BR_{D85N} but the net proton transport across the membrane is null. As a result, there is no formation of an M state and the states thereafter. Hence, the photocycle of a blue membrane differs substantially from a wild type. To recap, a blue membrane BR can be produced either by chemical treatment, removing or displacing divalent cations from a purple membrane, or by genetic mutation by replacing aspartic acid D85 with asparagine. In either case, the shelf lifetime by photoinduced refractive index modulation is stable for several months.

4.2   Reversible Reaction Scheme

A reversible reaction scheme for the simplest photocycle of bacteriorhodopsin is the transition between B to M state and vice versa. This photocycle, the simplified version of

Fig. 4-1, can either be induced by a one-photon or two-photon linear process.[69] One-photon excitation is a phototransition process that depletes the B-state and populates the M-state simultaneously.   The excited molecules thermally relax after the primary photochemical product under this process. A two-photon linear process is two sequential one-photon steps. Under this sequential one-photon process, the first illumination depletes the B-state causing molecules to concentrate in the M-state, and the second photoexcitation induces a population inversion. There are several intermediate states in the photocycle but the M-state is the only state that provides a stable photochemical transition back to the B-state. Hence, state M is considered to be the dominant intermediate in a holographic gratings. Thus, the simplified BR photocycle for a first order reversible reaction scheme under one-photon and/or sequential one-photon process is illustrated by Fig. 4-2 as follows:
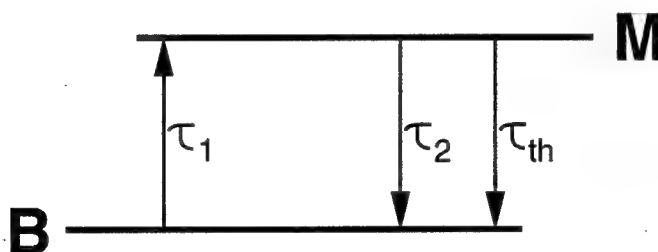
Fig. 4-2 The simplified bilevel photocycle of bacteriorhodopsin.

Time constants τ_1 and τ_2 are forward and reversed photochemical transformation from state B to M, and vice versa. The constant τ_{th} is the thermal relaxation time. Let's put these time constants in a proper perspective. Thermal lifetime τ_{th} is longer than time constants τ_1 and τ_2. Time constant τ_1 is the transformation time from state B to M via intermediates J, K,

and L. This constant is longer than τ_2 since τ_2 is the direct population inversion time from M to B-state. The functional representations are $\tau_{th} \gg \tau_1$, and $\tau_1 > \tau_2$.

Time constants are inversely proportional to the illumination intensity. The expression for a forward reaction time constant from B to M-state is $\tau_1 = 1/(a_1(\lambda_1) \ I(\lambda_1))$, and the reversed reaction time constant from M to B-state is $\tau_2 = 1/(a_2(\lambda_2) \ I(\lambda_2))$. The reaction constants $a_1(\lambda_1)$ and $a_2(\lambda_2)$ are defined as $2.3026(\varepsilon_{B,M}\Phi_{B,M}\lambda_{B,M}/N_a ch)$, where the subscripts B,M are referred to that given parameter in either B or M-state.[66] The symbol $\varepsilon_{B,M}$ is the molar absorption coefficient, $\Phi_{B,M}$ is the quantum yield coefficient for forward/reverse reaction, N_a is Avogadro's number, h is Plank's constant, and c and $\lambda_{B,M}$ are the speed of light and the wavelength of maximum absorption at B and M-state, respectively.

4.3 Transition Rate by One-Photon Process

Photoexcitation with yellow illumination or thereabouts bleaches a BR sample causing its molecules to undergo a complex phototransformation. The effect of this illumination depletes the population from B-state and builds up molar concentration in the M-state. The next two subsections will explore the depletion rate of B-state and growth rate of M-state by one-photon process.

4.3.1 Depletion Rate of B-State

The difference between the depletion rate and growth rate of molar concentration in a given state is known as the rate equation. Under a one-photon process, B-state is

depleted by phototransformation, and it regains its molar concentration via thermal relaxation. The rate equation for this state, by definition, is the rate of change of molar concentration with respect to time.[87] Hence, the time dependent kinetic rate equation of the B-state is :

$$\frac{dB(t)}{dt} = \frac{M(t)}{\tau_{th}} - \frac{B(t)}{\tau_1} \qquad (4\text{-}1)$$

where $B(t)$ and $M(t)$ are the population density of B and M-state as a function of time t, respectively. The total population B_0 is the molar concentration in the B-state under an initial light adaptive state. This population depends on material properties. Note that the value of this population is always constant. In other words, a total summation of BR molecules in both states, B and M, always equals B_0. Thus, the relation representing molar concentration between these two states is[57] :

$$B(t) + M(t) = B_0 \qquad (4\text{-}2)$$

All molecules initially concentrate in B-state. Hence, the initial M-state is completely unpopulated. Depleting BR molecules from the B-state increases molar concentration in M-state and vice versa. Note that a BR molecule relaxing into B-state can instantaneously excite into a new photocycle. Therefore, the initial conditions are :

$$M(t) = 0 \quad \text{at } t = 0 \qquad (4\text{-}3)$$

$$B(t) = B_0 \quad \text{at } t = 0 \qquad (4\text{-}4)$$

Substituting Equ. (4-2) into Equ. (4-1) yields the steady state rate equation as follows:

$$\frac{dB(t)}{dt} = \frac{B_0}{\tau_{th}} - \frac{B(t)}{\tau}; \quad \text{where} \quad \frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_{th}} \qquad (4\text{-}5)$$

The general solution of a nonhomogeneous linear differential equation of order 1 with constant coefficients is composed of two parts: a homogeneous solution and a particular solution. The homogeneous solution of Equ. (4-5) can be determined by:

85

$$\frac{dB(t)}{dt} = -\frac{B(t)}{\tau} \qquad (4\text{-}6)$$

The solution to Equ. (4-6) is given by:

$$B_h(t) = c_2 e^{-\frac{t}{\tau}} \qquad (4\text{-}7)$$

where c_2 is an arbitrary constant. The particular solution to Equ. (4-5) is found to be:

$$B_p = \frac{\tau}{\tau_{th}} B_0 \qquad (4\text{-}8)$$

Therefore, the general solution to Equ. (4-5) is a combination of $B_h(t)$ and $B_p(t)$. The solution becomes:

$$B(t) = B_p(t) + B_h(t) = \frac{\tau}{\tau_{th}} B_0 + c_2 e^{-\frac{t}{\tau}} \qquad (4\text{-}9)$$

An arbitrary constant c_2 can be determined by invoking the initial condition given by Equ. (4-4). The result becomes $c_2 = B_o (1 - \tau/\tau_{th})$. Thus, the solution to the initial value problem is:

$$B(t) = B_0 e^{-\frac{t}{\tau}} + \frac{\tau}{\tau_{th}} B_0 \left(1 - e^{-\frac{t}{\tau}} \right) \qquad (4\text{-}10)$$

The constant multiplication factor of the second term in Equ. (4-10) is referred to as the population decay of M-state. Let's designate it as symbol m_1. Its functional expression in terms of τ_1 and τ_{th} can be written as follows :

$$m_1 = \frac{\tau}{\tau_{th}} B_o = \frac{1}{1 + \frac{\tau_{th}}{\tau_1}} B_0 \qquad (4\text{-}11)$$

This equation illustrates that rebuilding of B-state always begins with a population that is less than its initial concentration since m_1 is always less B_o by (τ/τ_{th})-factor. Hence, this

86

factor has a significant influence on growth rate of B-state. In short, molar concentrations in B-state can never achieve its initial population under a steady state illumination. Note that this equation becomes B_0 when time t equals zero. Conversely, it reduces to m_1 as time t approaches infinity ($t \to \infty$) under steady illumination. Furthermore, m_1 is an asymptotic value of plots a and b as shown in the next figure.

4.3.2 Growth Rate of M-State

The intermediate M is an excited state, and its population consists of excited molecules. Molar concentrations in the M-state grow proportionally to the depletion rate of the B-state. Under the one-photon-process, a reversible reaction can only be achieved by thermal relaxation. The growth rate of M-state can be represented by the time dependent kinetic rate equation as follows :

$$\frac{dM(t)}{dt} = \frac{B(t)}{\tau_1} - \frac{M(t)}{\tau_{th}} \qquad (4\text{-}12)$$

This equation is realizable if and only if it contains one dependent variable. The dependent variable B(t) can be eliminated by substitution of Equ. (4-2) into Equ. (4-12). The result becomes :

$$\frac{dM(t)}{dt} = \frac{B_0}{\tau_1} - \frac{M(t)}{\tau} ; \qquad \text{where} \quad \frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_{th}} \qquad (4\text{-}13)$$

The general solution of a nonhomogeneous first-order linear differential equation is a combination of a homogeneous and a particular solution. The solution to the homogeneous equation is :

$$M_h(t) = ce^{-\frac{t}{\tau}} \qquad (4\text{-}14)$$

where c is an arbitrary constant. The particular solution to the equation becomes :

$$M_p = \frac{\tau}{\tau_1} B_0 \tag{4-15}$$

The combination of equations (4-14) and (4-15) gives rise to the general solution of Equ. (4-13) as follows :

$$M(t) = \frac{\tau}{\tau_1} B_0 + c e^{-\frac{t}{\tau}} \tag{4-16}$$

Constant c can be determined by utilizing the initial condition given by Equ. (4-3). Hence, the solution of the initial value problem is :

$$M(t) = \frac{\tau}{\tau_1} B_0 (1 - e^{-\frac{t}{\tau}}) \tag{4-17}$$

A direct substitution of Equ. (4-10) into Equ. (4-2) yields the same result as given in Equ. (4-17). In addition, the summation of equations (4-10) and (4-17) always equals B_o. The constant multiplication factor of Equ. (4-17) is referred to as the population growth (m_2) of M-state. It is inversely proportional to m_1. Furthermore, Equ. (4-17) approaches m_2 under long steady illumination ($t \to \infty$). Expressing m_2 in a functional form will help bring its definition to the surface. The simplified expression becomes :

$$m_2 = \frac{1}{1 + \frac{\tau_1}{\tau_{th}}} B_o \tag{4-18}$$

The significance of this equation is mainly to demonstrate that molar concentration in the M-state can increase its population up to B_o for those samples with a long thermal lifetime.

Molecule distribution between B and M-state a under one-photon process, as shown in Fig. 4-3, can be further studied by computer simulation of equations (4-10) and (4-17). They are referred to as B-state depletion rate and M-state growth rate equation, respectively. The simulation of Equ. (4-10) produced plots a and b. Likewise, the simulation of Equ. (4-17) produced plots c and d. The photoexcitation and thermal

88

relaxation transforms molecules back and forth between both states. This transformation is governed by a molar conservation law. For instance, a summation at each time distribution of plots a and c, or plots b and d always equal to B_o. Hence, these two equations are referred to as molar transformation pair.
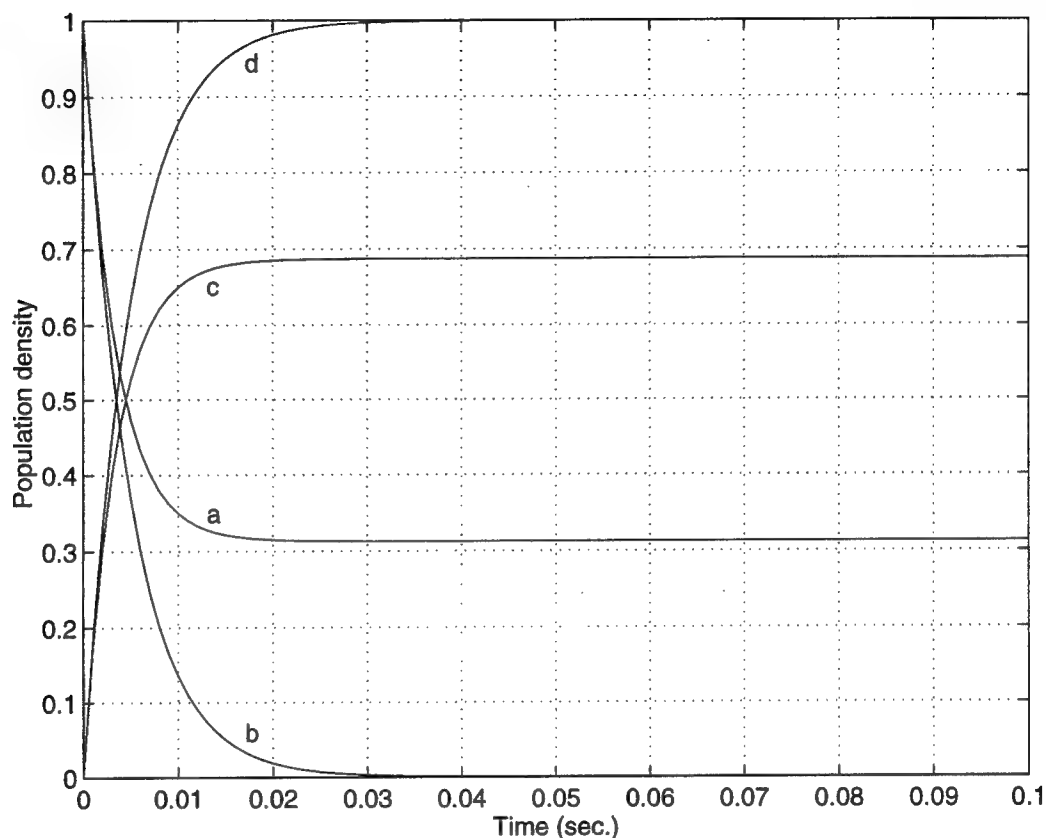
Fig. 4-3 Distribution of BR molecules in both states, B and M, by one-photon process. Thermal lifetime : $\tau_{th} = 11$ msec. for plots a and c; $\tau_{th} = 10$ sec. for plots b and d.

The parameters used in the simulations were $B_0 = 1$, $\tau_1 = 5$ msec., and $\tau_{th} = 11$ msec. for plots a and c, and $\tau_{th} = 10$ sec. for plots a and d. The value of the population density B_0 is normalized to unity. The time constant value of forward photochemical reactions was chosen to demonstrate the separation between plots a and b, and c and d.

The relaxation time constant of plots a and c is the thermal lifetime of the wild type BR,[88] and the relaxation value of plots b and d is a minimum M-state thermal lifetime of BR_{D96N}. It is evident from the figure that molar concentrations in M-state have potential to gain all BR molecules while B-state population can be completely depleted by strong illumination for those samples with a long thermal lifetime. On the other hand, both states only retain partial molecule distribution if the thermal lifetime of the sample is not long enough.

4.4 Transition Rate by Two-Photon Sequential Process

The two-photon sequential (linear) process is a process occurring in two sequential one-photon steps. An illumination in the yellow region transforms molecules from the B-state to an excited M-state. During the lifetime of the M-state, a second photoexcitation from a blue beam around 413 nm inverts the population in the excited M-state back to the B-state. Since the B and M-state have two distinctive spectral absorption bands, an interaction of these light beams in the BR film is an uncorrelated process. This reversible reaction scheme, as shown in Fig. 4-4, demonstrates a one-photon transition and a two-photon transition as shown in part a and b, respectively. The physics of a sequential one-photon process make it possible for this population inversion.

$$B \xrightarrow{\tau_1} M \qquad\qquad M \underset{\tau_1}{\overset{\tau_2,\tau_{th}}{\rightleftarrows}} B$$

a   b

Fig. 4-4   The reversible photoreaction model by two-photon sequential process.

a.) Forward photoreaction scheme by one-photon transition.

b.) Reversible photoreaction scheme under two-photon illumination.

The forward photochemical reaction, as shown in Fig. 4-4a, transforms molecules from the B to the M-state. This phototransformation will deplete the B-state. Thus, the equation to represent the depletion rate is :

$$\frac{dB(t)}{dt} = -\frac{B(t)}{\tau_1} \tag{4-19}$$

Applying the initial condition of Equ. (4-4) to the solution from Equ. (4-19) yields the result as follows:

$$B(t) = B_0 e^{-\frac{t}{\tau_1}} \tag{4-20}$$

The relation between the B and M-state is governed by a duality principle. The duality principle maintains that one state must lose its population in order for other state to gain its molar concentration. Thus, the depletion of the B-state causes molar concentration in the M-state to grow. Hence, substitution of Equ. (4-20) into Equ. (4-2) yields the functional expression for the growth rate of the M-state as follows:

$$M_b(t) = B_0 (1 - e^{-\frac{t}{\tau_1}}) \tag{4-21}$$

Note that this equation describes only the growth rate of the M-state. But excited molecules will soon thermally relax back to the B-state. Therefore, the lag time between first and second illumination must not exceed the M-state thermal lifetime. If it did, a phototransition that inverts population from the M to B-state will have low efficiency.

A second illumination at a wavelength around 410 nm, as shown in Fig. 4-4b, will induce a population inversion. In addition to the photoreversible reaction, molecules may also thermally relax from the M-state. To incorporate both reactions, the time dependent kinetic rate equation of the M-state can only be expressed in this form:

$$\frac{dM(t)}{dt} = \frac{B(t)}{\tau'_1} - \left(\frac{1}{\tau_2} + \frac{1}{\tau_{th}} \right) M(t) \tag{4-22}$$

91

This equation has two dependent variables. However, the dependent variable B(t) can be eliminated from the equation by substituting Equ. (4-2) into Equ. (4-22). The result becomes:

$$\frac{dM(t)}{dt} = \frac{B_0}{\tau'_1} - \frac{M(t)}{\tau}; \quad \text{where} \quad \frac{1}{\tau} = \frac{1}{\tau'_1} + \frac{1}{\tau_2} + \frac{1}{\tau_{th}} \tag{4-23}$$

This equation is a nonhomogeneous first-order linear differential equation. Its general solution is a composition of a homogeneous and a particular solution. Thus, the general solution to Equ. (4-23) becomes:

$$M_d(t) = M_p + M_h = \frac{\tau}{\tau'_1} B_0 + ce^{-\frac{t}{\tau}} \tag{4-24}$$

BR molecules initially concentrate in the B-state. Hence, it gives rise to the initial condition as given by Equ. (4-3). Applying this initial condition to Equ. (4-24) yields the solution for a constant multiplier c (c = - $(\tau/\tau_1)B_0$). Therefore, the solution to Equ. (4-24) becomes:

$$M_d(t) = \frac{\tau}{\tau'_1} B_0 (1 - e^{-\frac{t}{\tau}}) \tag{4-25}$$

Under a two-photon linear process, two laser beams at two different wavelengths must be used to excite molecules in a BR sample. One laser beam is used to build up molar concentration in the M-state, and the use of a second laser beam will induce a population inversion. The growth and decay rate of BR molecules in the M-state depend on the intensity of laser beams 1 and 2, respectively. Hence, the net molar concentration in the M-state can be written as follows:

$$M(t) = M_b(t) - M_d(t) = B_0(1 - e^{-\frac{t}{\tau_1}}) - \frac{\tau}{\tau'_1} B_0(1 - e^{-\frac{t}{\tau}}) \tag{4-26}$$

Both τ_1 and τ'_1 are forward phototransition time constants. The constant τ_1 is the photoreaction conversion time with B-state absorption illumination from a single laser

source. The forward reaction time with both illuminations, B and M-state absorption laser beams, is denoted by τ'_1. Note that both constants will eventually settle down to a constant reaction time. Under a steady state condition, both forward time constants are indistinguishable ($\tau_1 = \tau'_1$).

The steady state response of the molar distribution in the M-state, as shown in Fig. 4-5, is generated by the use of Equ. (4-26) when both forward photochemical reaction time constants are the same. In other words, the ratio of the M-state's build-up to depletion is convergent. For example, if the forward time constant has two different values then the distribution is the product of two different ratios. For simplicity, the population density B_0 is normalized to unity. The computer simulations are divided into three groups. The parameters used in group a were $\tau_1 = 50$ μsec., $\tau_{th} = 10$ msec., and $\tau_2 = 0.1$ μsec for plot a1, $\tau_2 = 19.5$ μsec for plot a2 and $\tau_2 = 65$ μsec for plot a3. Groups b and c have a thermal lifetime of 10 sec. ($\tau_{Th} = 10$ sec.). The forward photoreaction time constant (τ_1) of group b was 0.5 sec., and the reversed photoreaction time constants were $\tau_2 = 1$ msec. for plot b1, $\tau_2 = 0.2$ sec. for plot b2 and $\tau_2 = 0.7$ sec. for plot b3. The values of the photoreaction time constants in group c were $\tau_1 = 1.0$ sec., and $\tau_2 = 1$ msec. for plot c1, $\tau_2 = 0.41$ sec. for plot c2 and $\tau_2 = 1.5$ sec. for plot c3. These simulations have shown that a large forward phototransformation time constant produces a slow rise time, and a slow reversed phototransition time constant deprives the molar concentration in the M-state from reaching its maximum potential. One of the reasons is that the slow reversed time constant allows more molecules to deplete from the M-state via thermal relaxation.
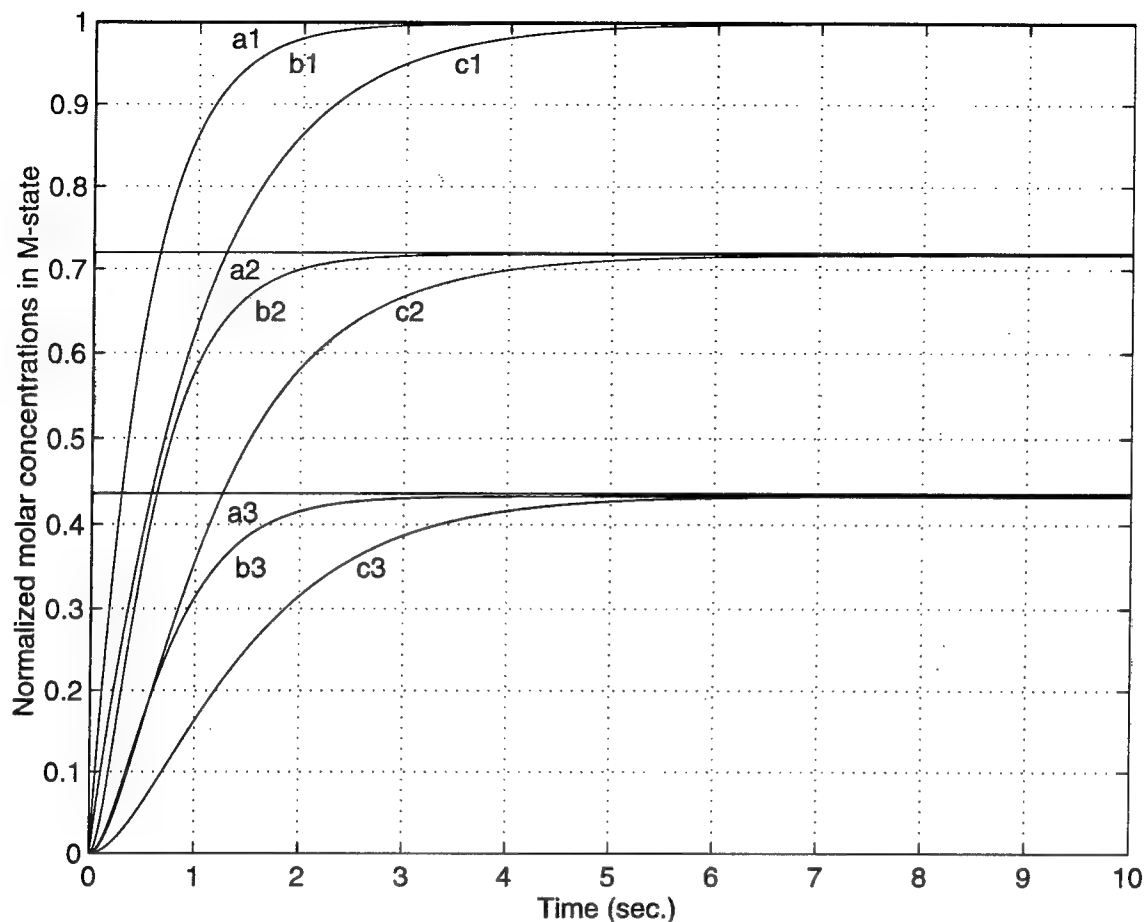
Fig. 4-5 Molar distribution in the M-state under a steady state two-photon excitation.

Let's further study molecule distribution in the M-state when a forward time constant has two different values. They are the products of two different phototransformation rates. Thus, Equ. (4-26) provides a good description for this type of molar distribution. The use of this equation produces results as shown in Fig. 4-6. The parameters of time constants used in the simulation were $\tau_1 = 0.2$ sec., $\tau'_1 = 4.0$ sec., $\tau_{th} = 10.0$ sec., and $\tau_2 = 2.0$ sec. for plot a, $\tau_2 = 3.0$ sec. for plot b, $\tau_2 = 4.0$ sec. for plot c, $\tau_2 = 5.0$ sec. for plot d and $\tau_2 = 6.0$ sec. for plot e. The values of time constants were chosen to emulate realistic behavior of molar distribution as if they were from an experimental measurement. Again, the value of population density B_o is normalized to unity. The

simulations have revealed that the graph first rises to a maximum peak before it decays to a steady state value. Each graph begins with a hump, then it flattens to an asymptotic level. The hump is a product of two different ratios. They are the ratio of the build-up to the depletion rate in the M-state. These ratios eventually converge to a constant rate forcing the graph to approach a steady value. Hence, the graph is called quasitransient. In addition, the hump spreads wider as the value of τ_1 increases, and raising the value of τ'_1 heightens the hump amplitude. In addition, increasing either the thermal lifetime τ_{th} or the reversed time constant τ_2 decreases the steady state level as shown in the figure.
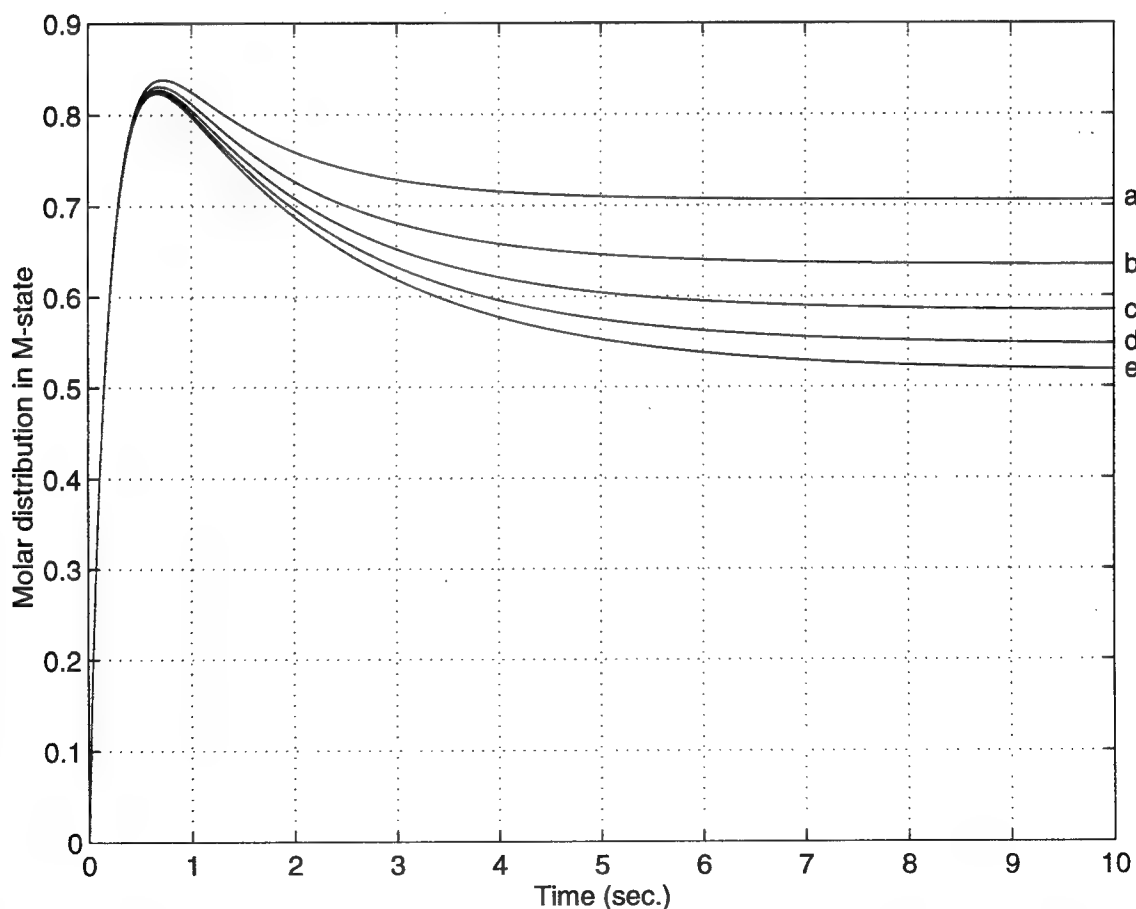
Fig. 4-6 Quasitransient molar distribution in the M-state by two-photon linear process.

4.5   Transient Dynamic Hologram

The transient dynamic hologram is a technique to investigate time dependent dynamics of the excited M-state in a BR sample. The optical generation of a transient configuration is the product of a sequential one-photon process. The use of the B-state absorption laser pumps BR molecules from the B-state (stable state) to the photoexcited M-state. However, excited molecules can not remain in the M-state exceeding their M-state thermal lifetime. It poses an interesting challenge to build a fully occupied M-state. Let us consider a realistic model that allows some molecules to thermally decay from an excited state after they have lived out their lifetime during a B-state depletion process as shown in Fig. 4-7a. Precise pumping duration and power are vital in maximizing molar concentration in the excited M-state. The M-state absorption laser is triggered open by the closing of a synchronous pump laser. Thus, the model for an induced phototransition from M to B-state by M-state absorption illumination is shown in Fig. 4-7b. Note that strong M-illumination eliminates any probability that some molecules still thermally decay during this process.

The rate equation governing a forward induced phototransition by B-state absorption illumination is:

$$\frac{dM(t)}{dt} = \frac{B(t)}{\tau_1} - \frac{M(t)}{\tau_{th}} \tag{4-27}$$

a   b

Fig. 4-7 Transient holographic model by a sequential one-photon process.

a.) Forward induced phototransition by B-state absorption illumination.

b.) Reversed induced phototransition by M-state absorption illumination.

It is not a coincidence that this equation is an exact duplication of Equ. (4-12). Both equations represent the growth rate of the M-state by a one-photon process. An important factor to be recognized is the M-thermal lifetime of a BR film should be long enough to trap most of excited molecules, or else a transient system will not produce maximum diffraction efficiency. Thus, the solution to Equ. (4-27) is given by Equ. (4-17) as follows:

$$M_1(t) = \frac{\tau}{\tau_{1}} B_0 (1 - e^{-\frac{t}{\tau}}) ; \qquad \text{where} \quad \frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_{th}} \qquad (4\text{-}28)$$

A complete build-up of the M-state population is the result of B-state absorption illumination on a BR film after Δt seconds. Then a synchronous M-state absorption laser is turned on by a triggering off a pump laser. In a simple model, the equation governing the M-state depletion rate after Δt-second pumping time is:

$$\frac{dM(t)}{dt} = -\frac{M(t)}{\tau_2} \qquad (4\text{-}29)$$

The physics of a second illumination is to deplete molar concentration in the M-state and establish a direct M->B phototransition. Hence, the solution to the M-state photodepletion rate is:

$$M_2(t') = c e^{-\frac{t'}{\tau_2}} \qquad (4\text{-}30)$$

The population in the excited M-state decays exponentially, and the decay rate depends on τ_2. It is a reversed phototransformation time constant created by the second illumination. Note that a second laser is turned on after Δt seconds. Consequently, the relation of the time sequence for turning on a laser between these two illuminations is given by $t' = t - \Delta t$. It is conventional wisdom to use zero time as an initial condition. Therefore, an initial condition of a second illumination is at $t' = 0$, implying that $t = \Delta t$. At the instance of

97

switch illumination, from B to M absorption laser, molecules in the excited M-state are depleted by the second laser. In short, the functional expression representing M-state molecules at this critical moment is $M_1(t=\Delta t) = M_2(t'=0)$. The initial condition of Equ. (4-30) is given by this relation. Hence, an arbitrary constant c can be determined by utilizing this relation. Therefore, the constant multiplier turns out to be $c = (\tau/\tau_1)B_0(1-\exp(-\Delta t/\tau))$, and the equation becomes:

$$M_2(t') = \frac{\tau}{\tau_1}B_0(1 - e^{-\frac{\Delta t}{\tau}})e^{-\frac{t'}{\tau_2}}$$
(4-31)

The transient profile is composed of two reactions: exponential rises and decays. The rise and decay rate depend on time constants τ_1 and τ_2, respectively. Note that the two functions exist sequentially in time. In other words, the time in which molecules were pumped to the M-state does not overlap with the time when they are photodepleted from the state. The functional description of a complete transient profile is a summation of time sequential counter reaction functions, $M(t) = M_1(t) + M_2(t')$. The mathematical expression can be written as follows:

$$M(t) = \frac{\tau}{\tau_1}B_0(1 - e^{-\frac{t}{\tau}})\Big|_{t=0}^{t=\Delta t} + \frac{\tau}{\tau_1}B_0(1 - e^{-\frac{\Delta t}{\tau}})e^{-\frac{t}{\tau_2}}\Big|_{t=\Delta t}^{t=\infty}$$
(4-32)

The computer simulation of Equ. (4-32) fully describes the graphic representation of the complete transient profile, as shown in Fig. 4-8, for a dynamic holographic material. The parameters utilized in the simulations were $\tau_1 = 0.1$ sec., $\tau_2 = 0.5$ sec., $\Delta t = 0.5$ sec., and $B_0 = 1$ which is the value of a normalized molecule population. There were no real scientific and/or physical considerations on the choices of these values. The figure displays the plots in descending order from a to g for thermal lifetime varying from $\tau_{th} = 10.0$ sec., 1.0 sec., 0.5 sec., 0.3 sec., 0.2 sec., 0.1 sec. and 0.05 sec., respectively. Plot a

corresponds to the longest thermal lifetime (τ_{th} = 10 sec.), and the shortest thermal lifetime

(τ_{th} = 0.05 sec.) produces plot g. This figure illustrates the fact that a BR sample with a

long M-state thermal lifetime incurs a strong population growth in the state, which leads to

high transient diffraction efficiency for the M-type hologram. It is important to recognize

the physics of the sample, because its thermal lifetime is constant at room temperature.

Hence, the various values of τ_{th} employed in the simulations represent various molar

concentrations in the excited M-state. Since fewer molecules remain in the M-state if it has

a short thermal lifetime τ_{th}, the transient system becomes inefficient.   A low

phototransformation power in an experimental application accomplishes the same task as a

short thermal lifetime in the simulation, because a low illumination intensity of the M-state

absorption laser allows more excited molecules to escape from M-state via thermal

relaxation.
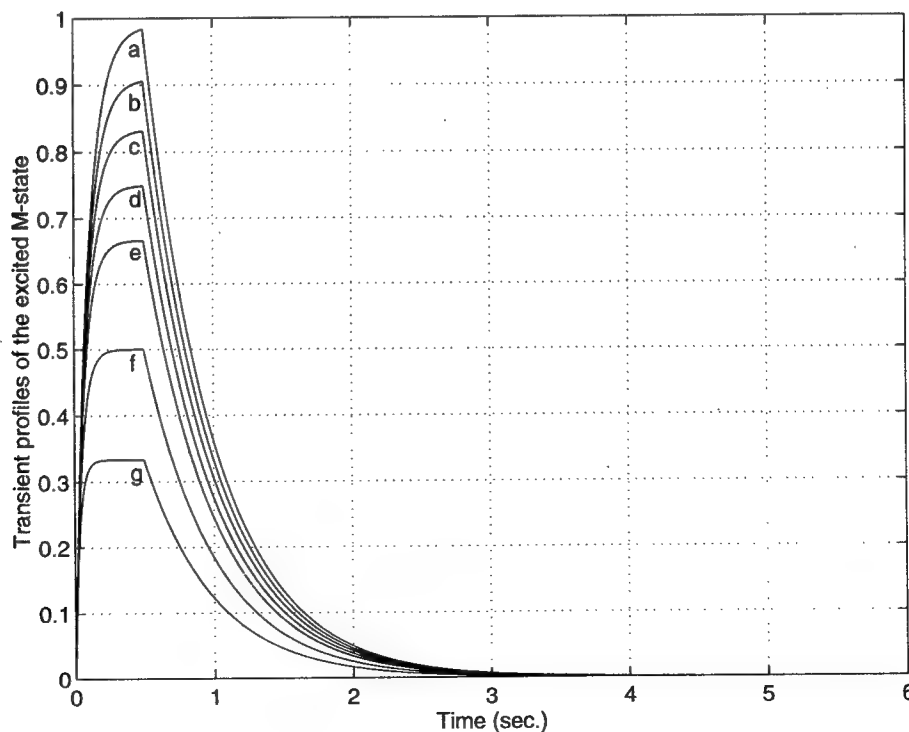
Fig. 4-8 Transient profile of a dynamic hologram.

4.6 Summary

Lengthy description of a bacteriorhodopsin and its various mutation variants is given at the beginning of this chapter. The model of reversible reaction scheme under one-photon process was investigated. The investigation is focused on depletion rate of B-state, which is proportional to the growth rate of M-state. A rate equation is used to represent this model. Hence, the model for a two-photon sequential process leads to a different rate equation, and the transient dynamic hologram has its own model and rate equation. The behavior of B-type holograms can be explained by a one-photon process, and a two-photon process is an explanation tool for M-type hologram and quasi transient. Hence, the theories developed in this chapter can be used to explain the observed behaviors of B-type and M-type holograms, and transient holographic gratings given in the next chapter.

Chapter 5

Holographic Gratings by Complex Photochemical Reactions

Holographic gratings produced by photochemical refractive index changes represent a method to evaluate diffraction efficiency and time dynamics of a BR based holographic process. Studying the underlying physics of a holographic recording and reading process can reveal optimum diffraction conditions in the time domain and intensity domain, and the best recording/reading configuration and wavelength. Since the refractive index modulation coefficient depends on both changes in the absorption coefficient and index of refraction, the photochemical gratings can be examined in wavelength regions where there is no overlap of reactant absorptions. Hence, a novel holographic technique is developed to investigate B-type and M-type holograms. The central issue of the experiment is to determine both maximum diffraction efficiencies and response times. The procedure has the virtue of experimental simplicity, and it is also applicable to transient grating measurements.

The objective of this chapter is to obtain experimental data by holographic gratings of complex photochemical reactions. The results can be explained by the theories developed in chapter 4. Three distinct setups used to obtain experimental data involve normal incidence of laser beams, B-type holographic measurements, M-type and transient holographic investigations. Section 5.1 provides descriptions of four BR films in terms of response time, absorption, reflection and transmission coefficients. Holographic diffraction efficiency and response times of B-type and M-type holograms are provided in Sec.5.2 and 5.3, respectively. Section 5.4 is devoted to transient holographic grating characteristics.

5.1   Normal Incidence of Laser Beams

The paraxial Helmholtz equation in a source-free lossy medium can be written in a homogeneous form as follows :

$$\nabla^2 E - 2ik \frac{\partial E}{\partial z} = 0 \tag{5-1}$$

where $\nabla^2 = \partial^2 / \partial x^2 + \partial^2 / \partial y^2$ is the transverse Laplacian operator. A Gaussian-beam solution of a transverse intensity distribution is everywhere a gaussian, and the solution to Equ. (5-1) becomes[29,30,89] :

$$E(x, y; z) = E_0 \frac{W_0}{W(z)} e^{-\frac{\rho^2}{W^2(z)}} e^{-i[(\frac{\rho^2}{2R(z)}+z)k-\Phi(z)]} \tag{5-2}$$

The solution is a composition of a Gaussian amplitude profile and a quadratic phase distribution , where k is the wave number of a Gaussian beam in a medium, W_0 is the beam waist radius (minimum spot size), which is a value of W at the plane $z = 0$, and E_0 is the field amplitude at the waist position. The 1/e radius of the field distribution is :

$$W(z) = W_0 \left[1 + (\frac{z}{z_0})^2 \right]^{1/2} \tag{5-3}$$

and the radius of curvature of an optical wavefront is represented by :

$$R(z) = z \left[1 + (\frac{z_0}{z})^2 \right] \tag{5-4}$$

where the spatial variation phase of the wave is an expression of :

$$\Phi(z) = \tan^{-1}(\frac{z}{z_0}) \tag{5-5}$$

with $z_0 = \pi W_0^2 n_0 / \lambda$, and n_0 is the intrinsic refractive index of a medium.

Let's consider wave propagation in a light absorbent medium. Hence, the wave number k in the paraxial Helmholtz equation is complex[25, 26] :

$$k = \beta - i\alpha \qquad (5\text{-}6)$$

where β is a phase constant, and α represents an absorption coefficient. The parameter α is also known as an attenuation or extinction coefficient. If both β and α are positive quantities, then substitution of Equ. (5-6) into Equ. (5-2) yields an attenuation factor, $\exp[- (\rho^2 / 2R(z) + z)\alpha]$, which attenuates as z increases. On the other hand, the parameter α in certain media is negative so that the substitution produces an amplification factor instead of attenuation. Note that $\alpha = 0$ for a lossless medium, and the wave number becomes a real phase constant.

BR film is one of those materials that has a positive absorption quantity. Its intensity dependent exponential absorption coefficient is given by[90, 91] :

$$\alpha = N\sigma_1 \left[\frac{1 + 2\sigma_2 \tau I / h\nu}{1 + (\sigma_1 + \sigma_2)\tau I / h\nu} \right] = \alpha_0 - \frac{gI}{1 + I/I_s} \qquad (5\text{-}7)$$

where N is the active density of BR molecules, σ_1 and σ_2 are the wavelength dependent absorption cross sections for the nonradiative transitions B-to-M and M-to-B state, respectively, τ is the thermal relaxation time from M-to-B transition, I is an illumination intensity, h is Plank's constant, ν is a frequency. The saturation intensity I_s equals $h\nu/(\sigma_1 + \sigma_2)\tau$, and the parameter g equals $N\sigma_1(\sigma_1 - \sigma_2)\tau/h\nu$, which depends on the M-state thermal lifetime and the absorption cross sections. The material absorption coefficient, $\alpha_0 = N\sigma_1$, only depends on the wavelength of the light beam. This equation illustrates the property of a BR film in which its absorption coefficient α decreases as laser intensity

103

increases. Thus, it seems to be senseless trying to provide a typical value for each constant in the previous equation because BR film can be prepared with different variants by chemical enhancement or genetic modification, and each variant will possess its own distinctive properties. Hence, the constants vary from variant to variant. However, the basic state configuration in the photocycle is not altered by the modification.

When a laser beam is incident on the surface of a medium that has higher refractive index, part of the beam reflects, some power dissipates in the medium, and the rest transmits through. Hence, three equations are needed to determine three unknown parameters of a laser beam propagation through a medium. These parameters are reflection coefficient (Γ), total absorption ($\forall$) and transmission coefficient (T). Let P_i be a laser beam incident power, P_r be the reflected power, and P_t be the transmitted power. Then the equations that describe the coefficients can be written as follows :

$$\Gamma = \frac{P_r}{P_i} \tag{5-8}$$

$$T = \frac{P_t}{P_i} \tag{5-9}$$

Since the summation of reflected, transmitted and dissipated power equals the incident power, then the total absorption in terms of reflection and transmission coefficient is given by :

$$\forall = 1 - (\Gamma + T) = \exp[- (\rho^2 / 2R(z) + z)\alpha] \qquad \text{(dimensionless)} \tag{5-10}$$

The absorption coefficient of a Gaussian beam propagation in a medium can be expressed in terms of the total absorption as follows :

$$\alpha = \frac{-\ln(\forall)}{\dfrac{\rho^2}{2R(z)} + z} \tag{5-11}$$

104

The absorption of a BR film is a function of wavelength and intensity. It absorbs strongly around the 570-nm wavelength for a B-type holographic application, and in M-type holograms, the strongest absorption is around 413 nm. Note that the refractive index changes of the films is relatively large compared to the absorption changes at wavelengths in the near infrared (670 - 800).[67, 70] Thus, phase sensitive applications require a wavelength from this region.

The constants of a laser beam normal incidence on four different BR variants are studied. The normal incident apparatus, as shown in Fig. 5-1, for the experimental measurement is composed of a 532-nm and a 680-nm light sources, a beam expander, an iris, a variable neutral density filter (VND), beamsplitter (BS), shutter (SH), detector heads (D), optical power meters (OPM), and an oscilloscope. The use of the beamsplitter enables experimental simplicity, where both input and transmitted power can be measured simultaneously. Two wavelengths were chosen to demonstrate the underlying physics of the amplitude and phase sensitive nature of the films. They were the 532-nm B-state absorption laser beam from a doubled Nd: YAG laser and the 680-nm phase-sensitive light source from a diode laser. Note that beams from both lasers were linearly polarized with the vertical polarization orientation. Four different BR variants, as shown in Table 5-1, were investigated in the experiment. The table provides a description of each film in terms of thickness, OD, pH, and degree of hydration. The investigations have revealed two sets of power transmission curves. There are four figures in a set, and each figure belongs to the transmission curves of each BR variant. One set of figures is a product of the 532-nm experiments and, the other set, results of the 680-nm experimental studies. The figures in each set have displayed a very striking similarity of power transmission behaviors with the figures in the same set. Hence, only one figure from each set was chosen at random to represent a typical power transmission characteristics. Both figures are from the same BR film. The film (BR1) used in the experiment was a wild type BR with a thickness of 200 μm, an O. D. of 5.5, pH of 9.0, and hydration less than 20%.
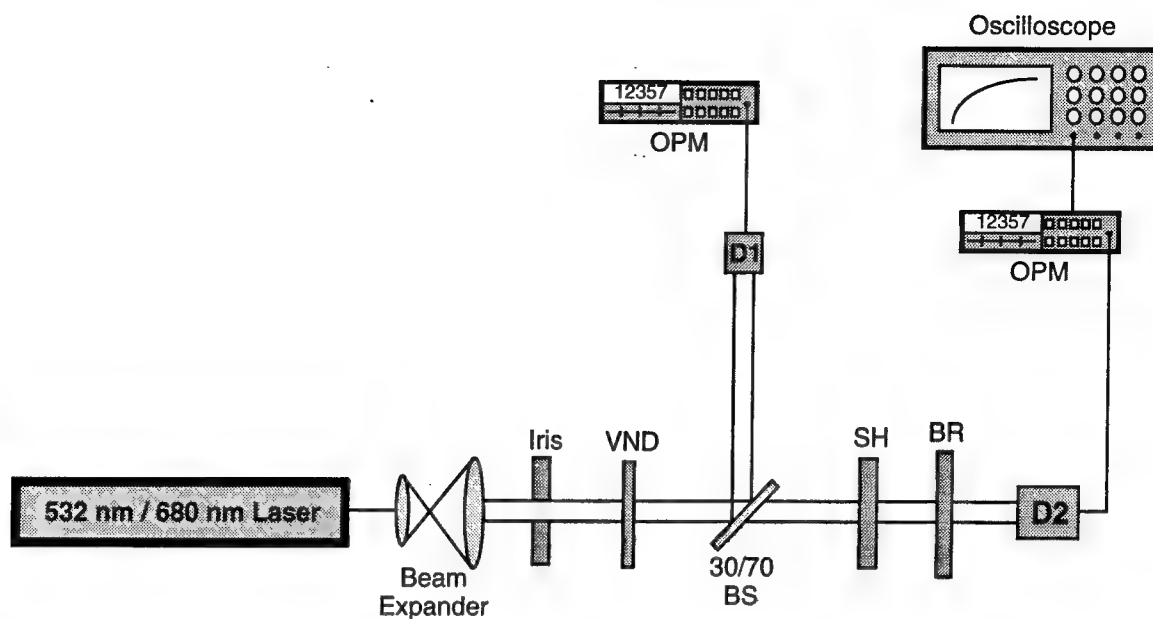
Fig. 5-1 Experimental setup to study BR's characteristics.

Film Label	BR's Type	Thickness	Optical Density	Solution pH	Degree of Hydration
BR1	BR_{WT}	200 μm	5.5	9.0	<20%
BR2	BR_{WT}	200 μm	5.5	9.0	>80%
BR3	BR_{D96N}	200 μm	6.0	9.0	<20%
BR4	BR_{D96N}	200 μm	6.0	9.0	>80%

Table 5-1 Description of BR films used in the experimental study.

Figure 5-2 provides an illustration of the 532-nm power transmission behavior of the BR1. The figure shows the behavior of the intensity-dependent transmission for various incident intensities, where a low intensity yields a slow response time. Note that a

strong incident power of a 532-nm laser beam quickly depletes molar concentrations from the B-state. The depletion rate is directly proportional to the illumination intensity. Therefore, the higher incident intensity produces a faster response time and larger transmission coefficient. This statement is fully supported and confirmed by the experimental data tabulated in table 5-2. Tables 5-2a and 5-2b are the experimental data of the wild-type and the D96N variant, respectively. The tables provide BR's characteristics in terms of incident intensity, response time, transmission coefficient, reflection coefficient and total absorption. The values of the coefficients and total absorption are determined by utilizing equations (5-8), (5-9) and (5-10), respectively. The reflection coefficients are independent of incident intensities because the reflections are mainly from both glass substrates that are sandwiching the BR-doped polymer film. The experimental results unambiguously disclose
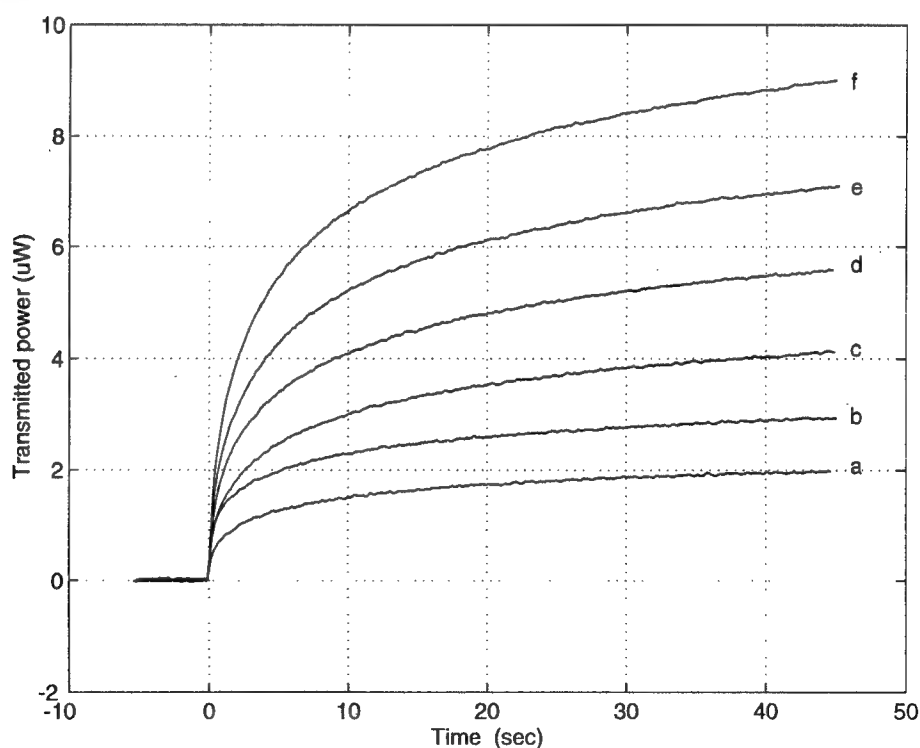
Fig. 5-2 Intensity-dependent transmission for various incident intensities. The intensities in mW/cm^2 are 49.44, 60.00, 70.39, 81.42, 89.90 and 100.64 for curves a, b, c, d, e and f, respectively.

107

The table provides a description of the film also in terms of response time, transmission coefficient, reflection coefficient and total absorption. The values of the last three parameters deviate from variant to variant as expected, but they are constant over a wide range of the incident intensities. The parameters are again determined by utilizing the same three equations. A conclusion from studying this data is that this wavelength can be used in a phase holographic application.

BR Film	Response Time	T	Γ	$\forall$
BR1	25.0 msec.	0.752	0.073	0.175
BR2	24.5 msec.	0.766	0.062	0.172
BR3	25.0 msec.	0.704	0.057	0.239
BR4	25.0 msec.	0.701	0.055	0.244

Table 5-3 Experimental characteristics of BR variants under 680-nm illumination.

5.2 B-type Hologram

An experimental setup for B-type holographic measurements, as shown in Fig. 5-4, consists of a 532-nm and a 680-nm light source, a beam expander, a variable neutral density filter (VND), mirrors (Mi where i = 1, 2, 3), a shutter (SH), 50/50 nonpolarizing beamsplitter cube (BS), BR film, red filter, detector heads (D), optical power meters (OPM), and an oscilloscope. A 532-nm beam from a doubled Nd: YAG laser is used to write holographic gratings in the B-state. The use of the beam expander in the setup is to collimate this beam to approximately 1.0-cm diameter. The collimated beam was then split

by the BS into two write beams which formed a Bragg angle as they were focused onto a BR film by mirrors M2 and M3. These two beams were carefully constructed within its coherence length. The writing beams were recombined to create a sinusoidal interference intensity grating in the film for a 9.95° Bragg angle. The angle was computed for a 1-μm grating period, and it was determined by utilizing Equ. (3-19), the Bragg equation. The refractive indices of the films used in the computation were 1.5395 and 1.5315 for a 532-nm and a 680-nm wavelength, respectively.[92] The fringes of a grating pattern are perpendicular to the film surface if the two writing beams are symmetrically incident with respect to the grating normal. The holograms were read out by a beam that was incident along its Bragg angle, which was 12.83° as determined by the same method used in the previous computation. The reading is done with the 680-nm light source from a diode laser, which is far away from the absorption bands. Readings with a wavelength in the phase hologram region could offer higher diffraction efficiency than amplitude holograms. The read beam was kept at a power level around 0.6 mW and a 0.6 cm diameter. Note that the holograms can be read with a low power level to ensure no erasing effect occurs during the process. However, the results may not reflect a realistic application if the power level is too low. It should also be noted that all laser beams were vertically polarized with respect to an optical table. A red filter was placed in back of the film to block out the writing beams. The objective of this experiment is to study holographic diffraction efficiencies and response times of four BR films listed in Table 5-1. The method used in the experiment was to keep the read beam at a constant power level and vary the power of the writing beams with the VND until an optimum efficiency was found. The VND was placed behind the shutter which was used to turn the writing beams on and off. The shutter was placed in a strategic location allowing time for the adjusted power to stabilize. Note that the power of the write beam was measured in front of the VND, with the reflection losses (from the BS

and the mirrors) and absorption loss (in the BS) factored out to determine a net writing power on the film.
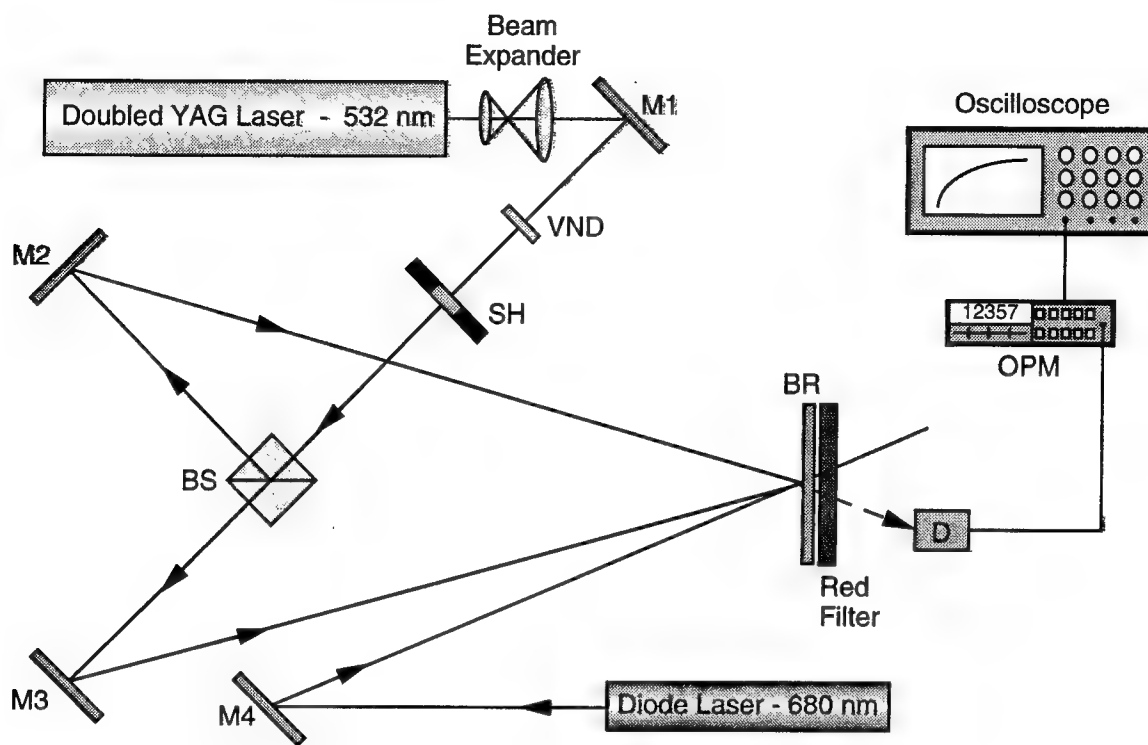
Fig. 5-4 Experimental setup for B-type holographic measurements.

Four curves, as shown in Fig. 5-5, are the diffraction behaviors in a Bragg regime of the films listed in Table 5-1. They increase with the incidence of the writing intensity and then flatten out to a plateau. The results seem to indicate that the grating structure was not degraded by the intensity of the read beam. These curves were produced by a similar writing intensity, and they were normalized to unity. Of course, the normalization does not intend to imply that they all have the same diffraction efficiency. Plotting them in this manner brings forth a comparison of their responses in real time. Hence, its true purpose is to illustrate the hologram rise time for four different samples. On the other hand, the curves of thermal decay as a function of time are shown in Fig. 5-6 which has the same labeling procedure as shown in the previous figure. In other words, alphabetical labels a,

b, c and d in both figures correspond to variants BR2, BR4, BR1 and BR3, respectively. The experimental data of grating decay behavior are depicted by solid line, and x-marks in the figure are the curve fit data which were generated by a computer simulation. Studying both figures leads to a conclusion that the hologram grating rise and decay time of a hydrated film is faster than a dry film, and response times of a wild type is quicker than D96N mutant. Note that thermal decay is considerably shorter than the M-state thermal lifetime for the BR_{D96N} which has a minimum lifetime in the excited M-state of approximately 10 seconds. However, the experimental data does not exceed 2 seconds for the longest decay. This discrepancy may, in part, be due to a low refractive index modulation coefficient.
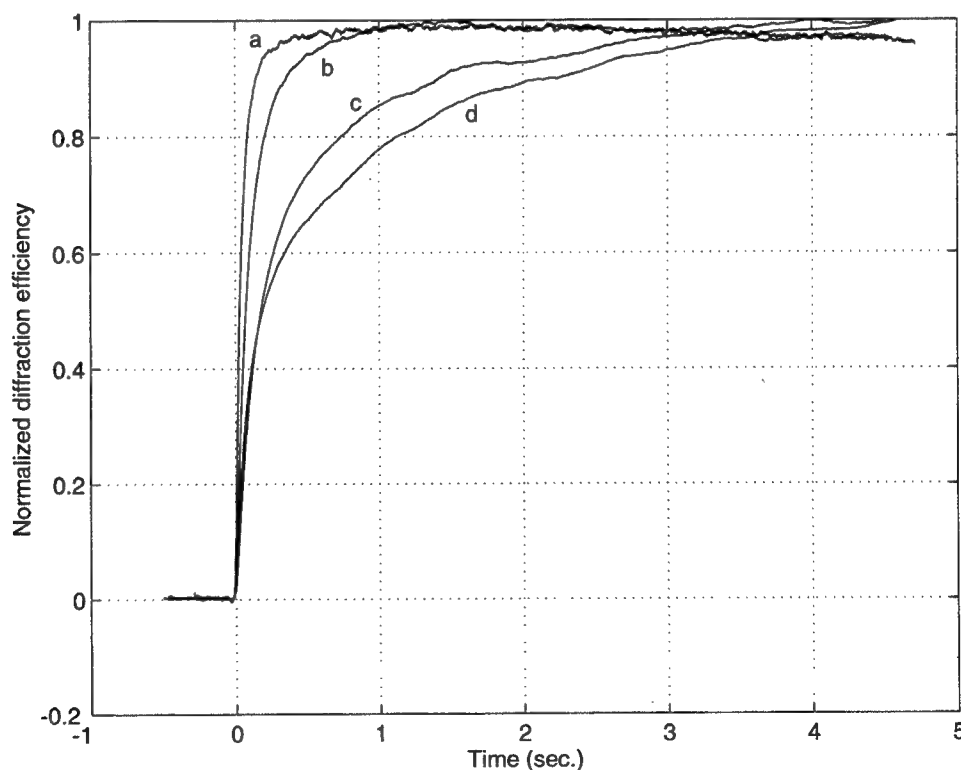

Fig. 5-5 Rise time behaviors of four BR films for a similar write intensity.

Plots a, b, c and d belong to BR2, BR4, BR1 and BR3, respectively.

113

It was found that the behavior of the thermal decay of the holographic gratings in all films could not be characterized by a single exponential function. The model that is best fit to the experimental data can be represented by a biexponential function as follows:

$$\tau_{th}(t) = P(a_1 e^{-(t/\tau_{1,th})} + a_2 e^{-(t/\tau_{2,th})}) \qquad (5\text{-}12)$$

where P is the scaling factor for the amplitude's adjustment, the summation of the amplitude constants a_1 and a_2 is unity (i. e., $a_1 + a_2 = 1$). This restriction is imposed for a normalized consideration. Time constants $\tau_{1,th}$ and $\tau_{2,th}$ are the fast and slow thermal lifetime, respectively. These lifetimes do not significantly depend on the write intensities. The biexponential decay can be explained by two distinct thermal lifetimes in the M-state. Thus, the results indicate that limited water content in a dry PM film not only slows down the responses, but also induces two distinct thermal lifetimes as shown in Table 5-4. The table provides a characteristic of each film in terms of write intensity, rise time, decay time,
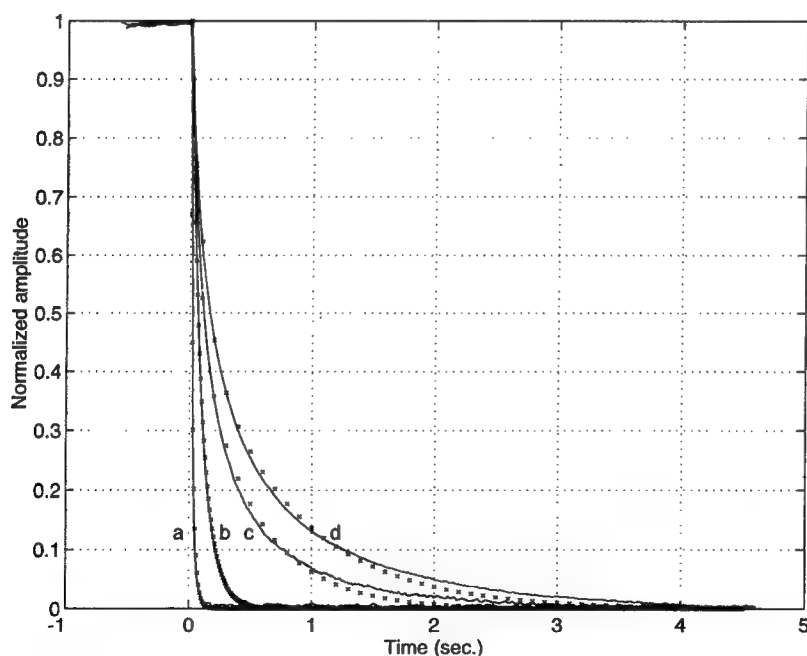
Fig. 5-6 Thermal decay behaviors as a function of time.

The experimental data are shown by solid lines and x-marks are curve fit data.

Plots a, b, c and d belong to BR2, BR4, BR1 and BR3, respectively.

and slow and fast thermal lifetimes. The rise time (t_{rise}) is defined as the time for the efficiency to increase from 10% to 90% of its maximum value. The definition can be expressed as follows :

$$t_{rise} = t_{90} - t_{10}$$ (5-13)

where t_{90} is the time at which the efficiency reaches 90% of the maximum value, and t_{10} is the time at which the efficiency is 10% of the maximum value. By the same token, the decay time (t_{decay}) is defined as the time for population gratings to decrease from 90% to 10% of its maximum value. The definition can be written as follows :

$$t_{decay} = t_{10} - t_{90}$$ (5-14)

These definitions are used throughout the chapter.

Film Label	I (mW/cm^2)	t_{rise} (sec.)	t_{decay} (sec.)	$\tau_{1,th}$ (sec.)	$\tau_{2,th}$ (sec.)
BR2	32.634	0.0905	0.051	0.025	0.025
BR4	29.488	0.361	0.263	0.095	0.095
BR1	34.157	1.350	0.799	0.070	0.480
BR3	30.471	2.019	1.535	0.100	0.770

Table 5-4 Holographic grating data by experimental measurement.

Figure 5-7 offers a different perspective on BR holographic gratings. It provides a description of the diffraction efficiency and rise time as a function of the write intensity. This figure plots out the curves of BR4. It was chosen to demonstrate the behavior of a BR film since the other three samples exhibit the same or similar behavior. The behavior

can be described as rising with the writing process and then flattening to a plateau. The plateau can be explained by the reading process. Reading with 680-nm laser beam does not erase nor alter the grating structure as the absorption of the film is virtually zero for this wavelength. Furthermore, the diffraction efficiency increases with the increase of writing intensity. A strong writing intensity produces high refractive index modulation coefficients, in which it yields high diffraction efficiency. In addition, grating fringes are
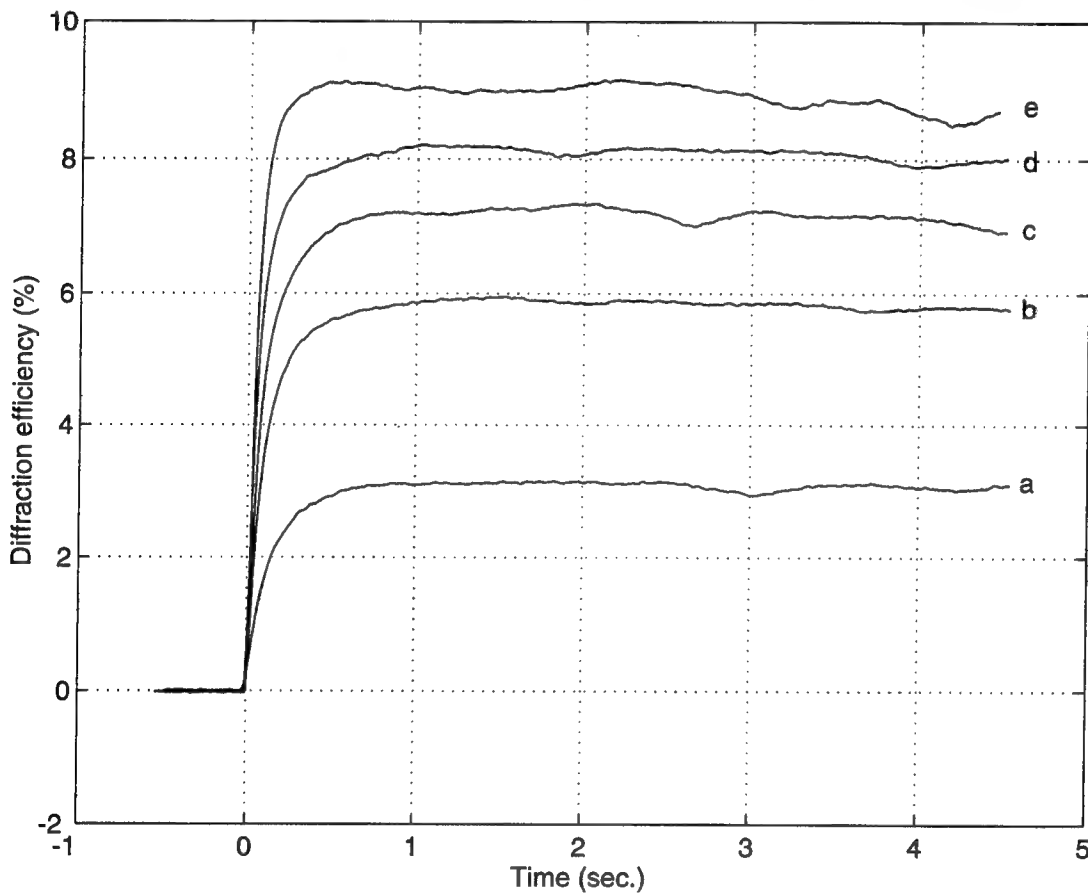
Fig 5-7 Diffraction efficiencies of the BR4 for various write intensities. The intensities in mW/cm^2 are 22.61, 53.37, 79.62, 108.12 and 172.02 for curves a, b, c, d and e, respectively.

formed quicker as both writing beams become more intense. This translates to a faster rise time. However, the gratings could saturate making them less sinusoidal if the writing intensity is too strong, resulting in a lower efficiency. It could also damage the film if the

writing intensity exceeds its optical limit. More explicit detail about the effect of a strong writing intensity will be discussed in the next two figures.

Figure 5-8 depicts the rise time of holographic gratings as a function of write intensity for all four BR films. The rise time for all films shortens with the increase of the writing intensity. A rising rate differs from sample to sample, nevertheless the curves reach their own flatness levels as the intensity increases. This conformity is hardly a surprise, because a forward photoconversion rate of all BR films is governed by the same
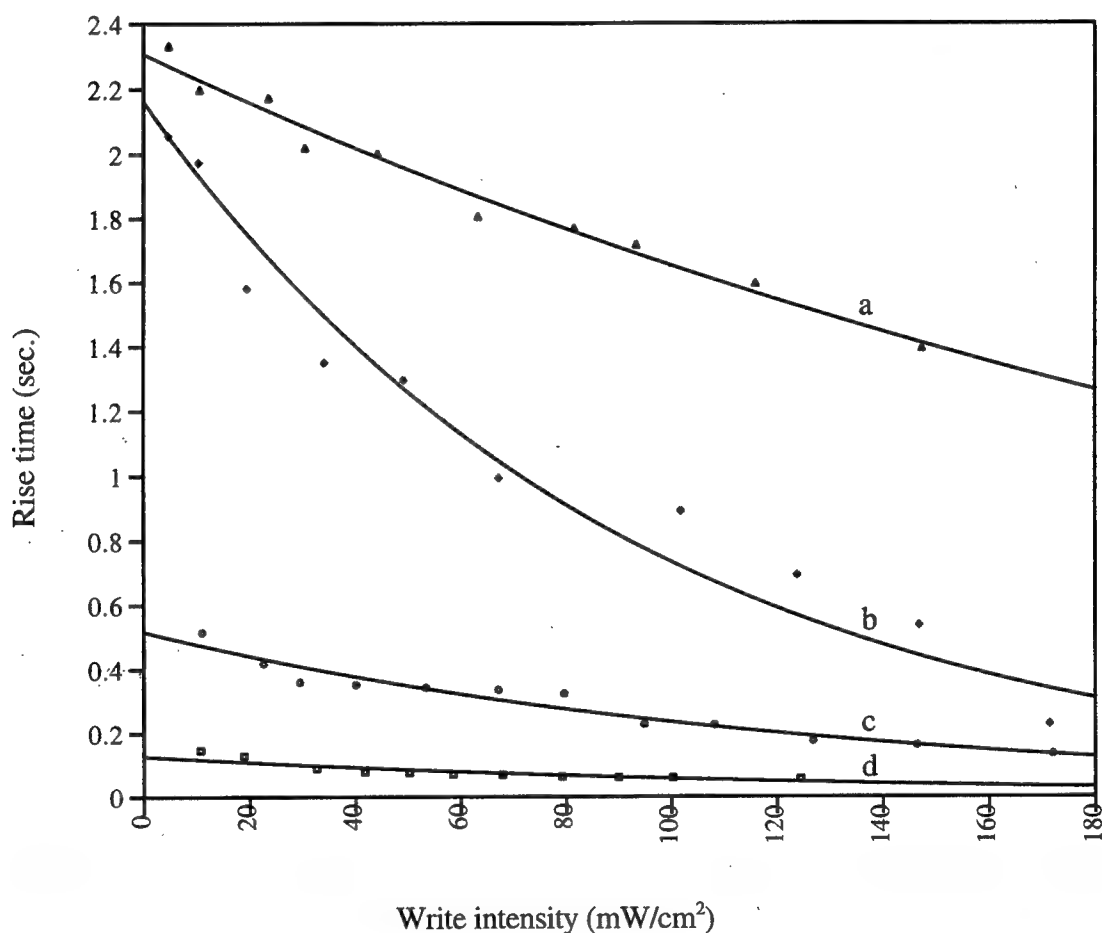
Fig. 5-8 Rise time of the holographic gratings as a function of write intensity. Plots a, b, c, and d correspond to the rise times of BR3, BR1, BR4 and BR2, respectively.

principle. The forward photoreaction time constant is inversely proportional to the writing intensity. Hence, a strong writing intensity produces a small time constant, which in turn,

117

yields a faster response in photoconversion from B-to-M. In other words, a fast rise time corresponds to a small time constant. Note that the reaction constant is a function of the film's characteristic. Therefore, the rate depends on the property of each film. For instance, the rise time of a hydrated BR film is shorter than a dehydrated one, and it is faster for a wild type BR film compared to a mutated D96N variant.

The experimental data of diffraction efficiencies as a function of write intensity for all four BR films are shown in Fig. 5-9. The efficiency is defined as the measured diffracted power (P_{dfr}) divided by the effective reading power (P_{eff}) which is defined as the incident reading beam power (P_{inc}) minus the glass surface reflected power (P_{rfl}). The functional expression for the efficiency is :

$$\eta = \frac{P_{dfr}}{P_{eff}} \tag{5-15}$$

and the effective reading power can be written as :

$$P_{eff} = P_{inc} - P_{rfl} \tag{5-16}$$

Note that each point on the graph is the peak value of the efficiency. All four films display a unified behavior, where their diffraction efficiencies increase proportionally to the writing intensity. The efficiencies approach their saturated values of 1.08%, 1.23%, 2.28% and 9.03% after the writing intensities exceed 171 mW/cm², 125 mW/cm², 147 mW/cm² and 172 mW/cm² for BR1, BR2, BR3 and BR4, respectively. Further, increasing the write intensity will not greatly optimize the efficiency to a new level, but the rise time seems to be faster. However, the efficiency could be compromised if the writing intensity saturates the holographic gratings. Note that the optical limiting of all films is greater than 180 mW/cm². Therefore, the gratings could be saturated long before the intensity exceeds a damaged threshold. The figure is favorable to the BR_{D96N} where its efficiency is higher than the BR_{WT}, and the efficiency of the hydrated BR film in both variants is greater than the dehydrated one. What's more, the results are unmistakably in favor of the BR4 because it has the highest efficiency and fairly short response times. Note that the M-state lifetime of

a BR_{D96N} is longer than a BR_{WT}. The long lifetime means low writing intensity is needed to achieve maximum grating amplitude, and fewer photons are required to keep BR molecules in the excited M-state. Therefore, BR film with a long M-state lifetime should have better sensitivity as given in the next figure. The reverse is also true for a BR film that has a short M-state lifetime.
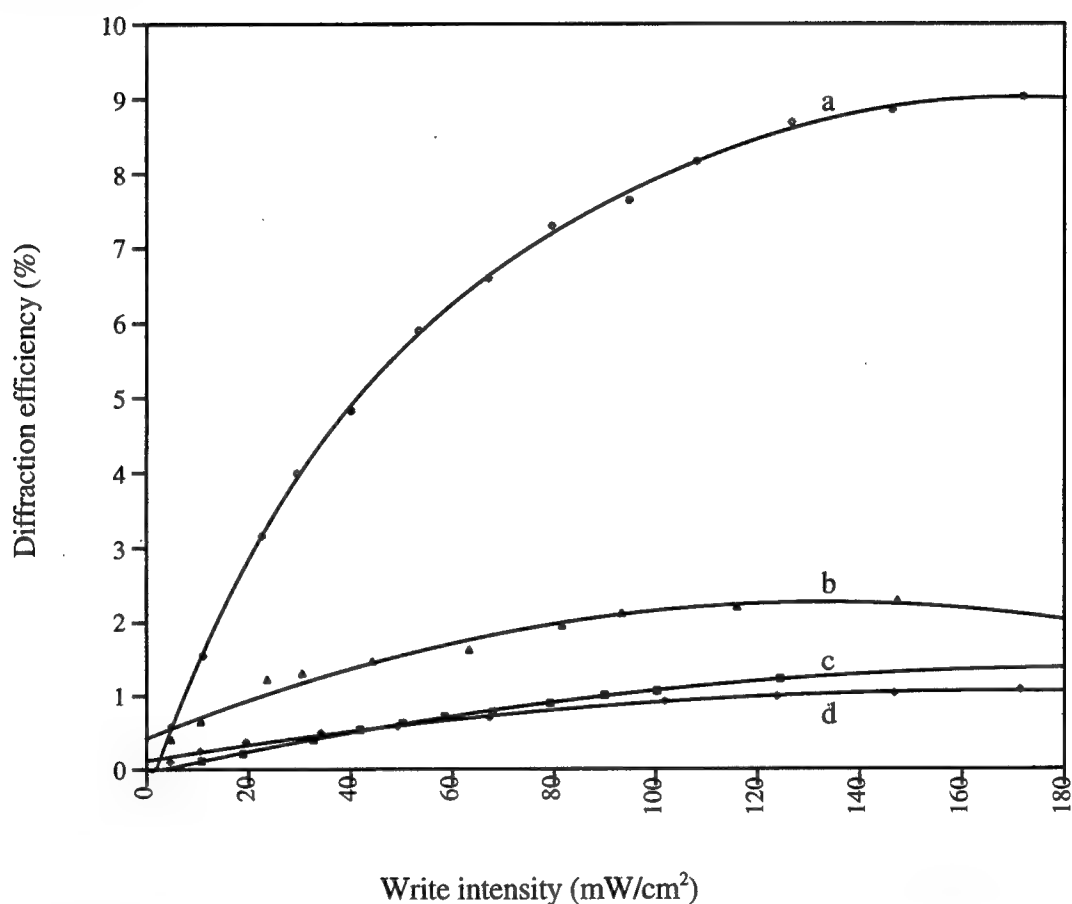
Fig. 5-9 Holographic diffraction efficiency as a function of write intensity. Plots a, b, c, and d correspond to the efficiency of BR4, BR3, BR2 and BR1, respectively.

Holographic sensitivities for all four BR films with respect to the 532-nm wavelength are illustrated in Fig. 5-10. In the figure, the commonly used reciprocal sensitivity S^{-1} is shown as a function of write intensity. A practical definition was applied

to determine the sensitivities. It is defined as the changes of diffraction efficiencies per incident energy density for unit thickness.[93] Since all samples have the same thickness, the unit of reciprocal sensitivity can be expressed in terms of millijoules per centimeters squared (mJ/cm^2). Then the definition can be written as follows: [69, 93]

$$S = \frac{\sqrt{\eta}}{I.t} \qquad (5\text{-}13)$$

where η is the diffraction efficiency, I is the write intensity, and t is the rise time for the efficiency to increase from 10% to 90% of its peak value. Note that the product of I.t is called exposure. The results illustrate that a hydrated film has better sensitivity than a dehydrated film, and a wild type film is more sensitive than a D96N mutant. Therefore,
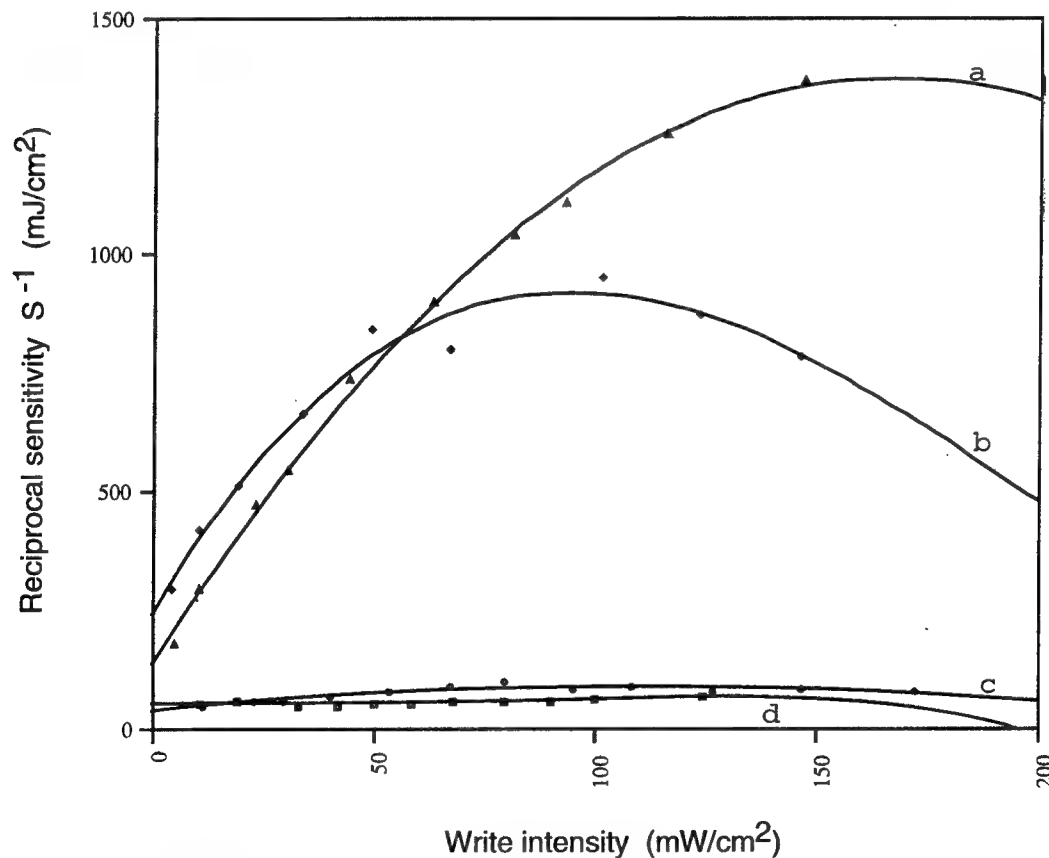

Fig. 5-10 Holographic grating sensitivity as a function of write intensity. Plots a, b, c, and d correspond to the sensitivity of BR3, BR1, BR4 and BR2, respectively.

the M-state lifetime of a hydrated film is longer than a dehydrated one. It should be noticed that these results are analogous to the plots of rise time shown in Fig. 5-8, where both sensitivity and rise time depend strongly on the write intensity. Hence, a better sensitivity would be found for all BR films at write wavelengths closer to 570-nm which is the B-state absorption maxima.

Cross talk, or channel interference, is a serious issue in designing a component for a communication network. The system provided in Fig. 2-6 will incur cross talk if and only if it operates in the Raman-Nath diffraction regime. The experimental setup shown in Fig. 5-4 was further modified to a 3-μm grating period, and the corresponding Bragg angles were 3.30° for the write beams and 4.24° for the read beam. On the other hand, Equ. 3-20 was employed to determine diffraction angles in the Raman-Nath regime. The angles were 4.25° and 12.83° with respect to the film normal of the back surface for a first and second order diffraction angle, respectively. It is customary to express the comparison unit in decibel (dB). This unit represents the power level of a second order diffraction (P_2) with respect to the first order diffraction power (P_1). The definition can be written as follows :

$$P_{dB} = 10 \log \frac{P_2}{P_1} \qquad (5\text{-}14)$$

The intensities were kept at 117.95 mW/cm^2 for the write beams and 3.98 mW/cm^2 for the read beam in the experiment. The second order diffraction powers were 21.53 dB, 21.66 dB, 23.17 dB and 27.42 dB below the first order diffraction powers for BR1, BR2, BR3 and BR4, respectively. Note that the second diffraction beams were barely discernible from the background. Therefore, the numbers above would be higher if background radiation can be eliminated from the measurements.

5.3   M-type Hologram

A three laser beam technique was developed to study M-type holograms with respect to some of their holographic characteristics. The method is to build up molar concentrations in the M-state before holographic gratings are written. The experimental setup to achieve this objective is given in Fig. 5-10. It consists of a 413-nm, 532-nm and 680-nm light source, two beam expanders, two shutters (SH1 & SH2), two variable neutral density filters (VND1 & VND2), six mirrors (Mi where i = 1, 2, 3, 4, 5 and 6), 50/50 nonpolarizing beamsplitter cube (BS), BR film, red filter, detector head (D), optical power meter (OPM) and an oscilloscope. A 532-nm light source from a doubled Nd: YAG laser is utilized to pump BR molecules from the B-state to the excited M-state, then the M-state absorption beam at 413-nm wavelength from a Krypton Ion laser is used to write holographic gratings in the M-state, hence it is called M-type hologram.[69] Note that the writing process drives some populations in the excited M-state back to the B-state. The purpose of the beam expander is to ensure uniform amplitude and constant phase across the beam diameter. The CW pump beam was collimated to an approximately a 1.0-cm diameter by the beam expander. This collimated beam was directed by mirrors M5 and M6 to be incident normal to the BR surface. The CW write beam was also collimated to about the same diameter as the pump beam. It was then split by the BS into two write beams which form a 7.65° Bragg angle between the write beam and the film normal as they were focused onto a BR film by mirrors M2 and M3. Equ. (3-19) was invoked to determine the Bragg angle for a 1-μm grating period. A refractive index of 1.551 for 413-nm wavelength was used in the computation.[92] Special care was taken to ensure that the recombination of the write beams was within its coherence length. The recombination creates a sinusoidal interference intensity grating at the BR film plane. Furthermore, slanted grating patterns with respect to the film surface can be prevented by maintaining symmetrical incidences of
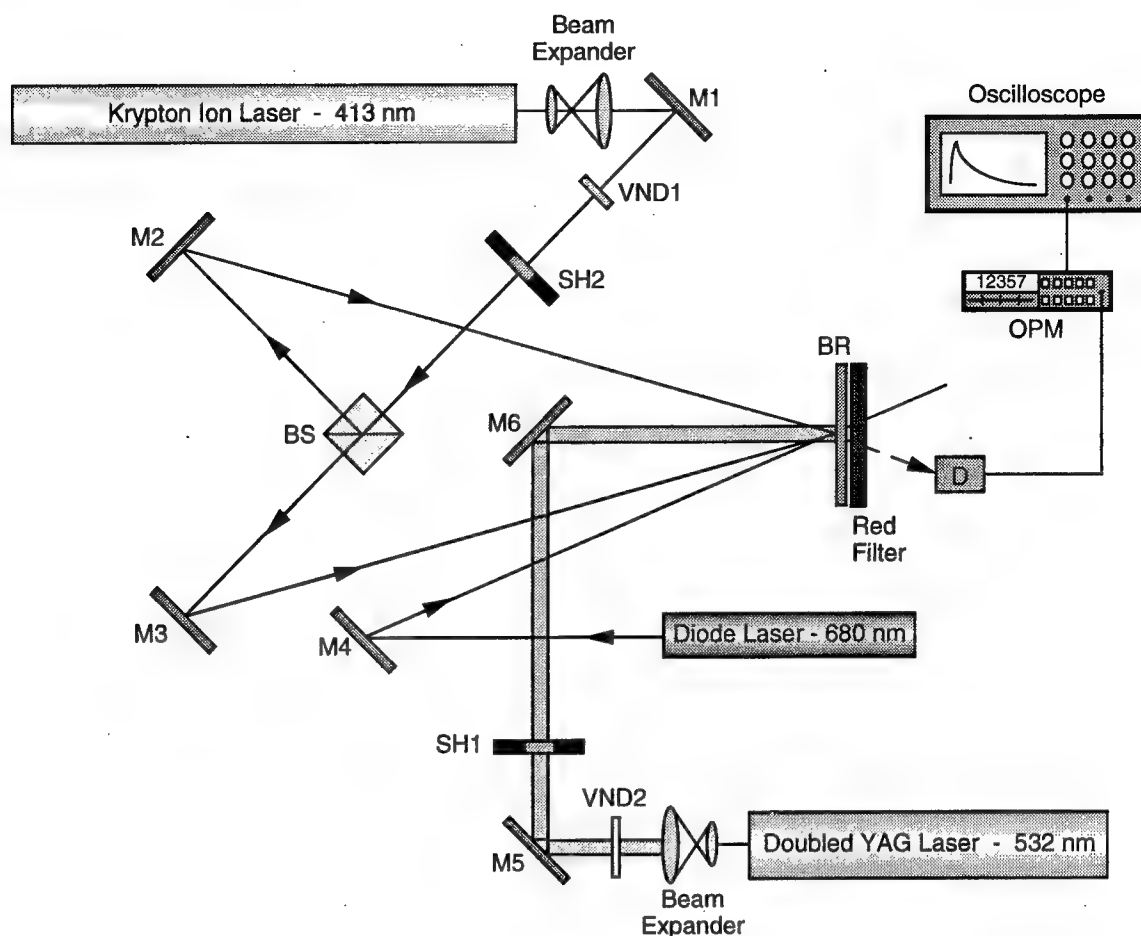
Fig. 5-11 Experimental setup for M-type holographic measurements.

the two write beams about the film normal. Shutters SH1 (controlling the 532-nm pump beam) and SH2 (controlling the 413-nm write beam) were synchronized such that SH2 was triggered open 2.5 seconds after SH1. Thereafter, both shutters remained open for an entire reading, which allowed the molecules to remain active in the photocycle between the B and M-state. Various pumping and writing conditions can be obtained by controlling the VNDs in the paths of pump and write beam. Measurements were taken behind the shutters after each power adjustment. A net writing power can be determined by considering the reflection losses of the mirrors and the absorption loss in the beamsplitter. In practice as well as in principle, the diameters of pump and write beam should be larger than the diameter of a read beam. The reading is done with the 680-nm phase-sensitive light source

from a diode laser. The read beam was approximately 0.6 cm in diameter and power level was attenuated to around 0.6 mW. It was incident along its Bragg angle which was 12.83° as determined by the same method used in the previous computation. Note that all three beams were vertically polarized with respect to an optical table. The purpose of the red filter behind the BR4 is to allow only the read beam pass through to the detector. Note that only BR4 was used in this experimental investigation. The measurements were read from the OPM, and an oscilloscope was used to record experimental data. The procedures are very much similar to the B-type hologram. In fact, a further objective of this experiment is also to investigate holographic diffraction efficiency and response times with respect to pump and write power by keeping the read beam at a constant power level and varying the ratio of write to pump beam until optimum results are found.

Figure 5-12 illustrates Bragg diffraction behavior of the M-type hologram as a function of pump intensity. The holographic gratings were created by a fixed write intensity at 31 mW/cm^2. The diffraction curves display a hump-like shape followed by an asymptotic tail at low pump powers. Hence, the behavior is called quasi transient. The behavior can be described as increasing efficiency to a peak value, at an optimal exposure time, followed by a decrease to a steady state value. The hump may be caused by a nonlinear saturation of a refractive-index modulation in a bilevel system. In other words, the rate of the population inversion may greatly exceed the rate of build-up in the M-state, which in turn, causes an initial overshoot in grating the structure. The amplitude of the hump increases and its width widens with the increase of pump power as shown in plots a and b. Thus, the quasi-transient plot reaches its highest efficiency at a pump intensity of approximately 13 mW/cm^2. Raising the pump power further will attenuate the efficiency, but a transition from quasi transient to a steady state begins to take shape. The first steady state curve, as shown by plot c, occurred at an intensity around 20 mW/cm^2. Plot d was the result of a stronger pump power saturating the holographic gratings, making them less

sinusoidal in structure. Therefore, the steady state diffraction efficiency decreases with the increase of pump intensity as shown in the next figure.
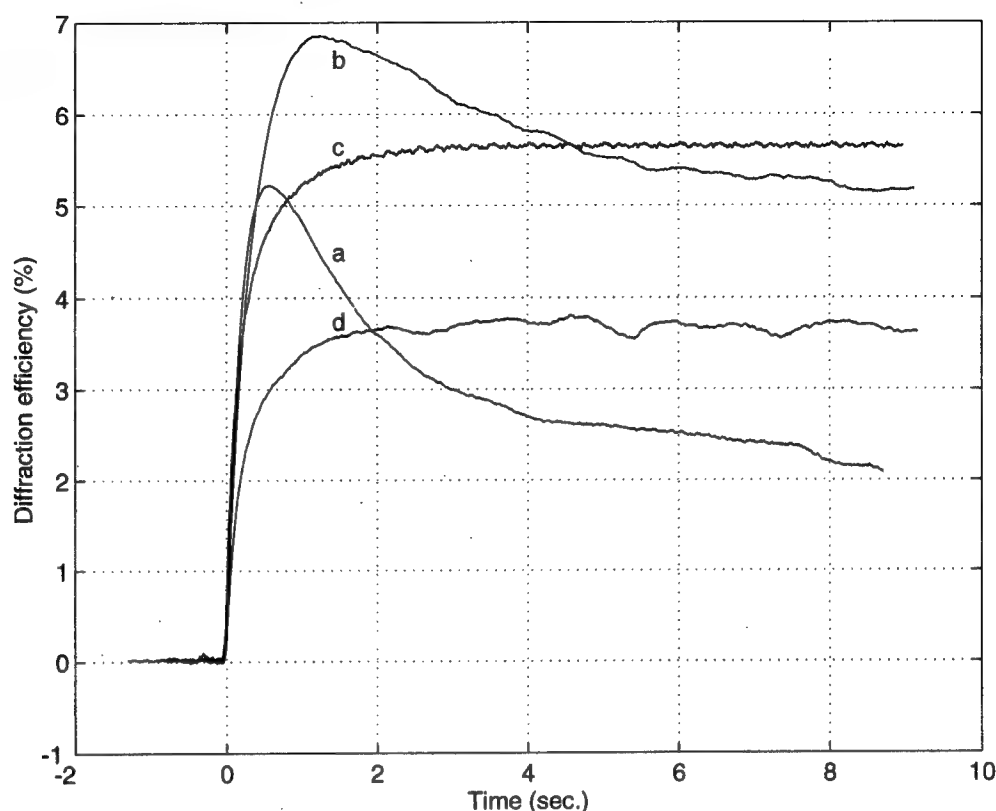
Fig. 5-12 Diffraction efficiency for various pump intensities. The intensities in mW/cm^2 are 5.10, 7.70, 19.51 and 28.53 for plots a, b, c and d, respectively.

Figure 5-13 shows the minimized Bragg diffraction efficiency as a function of pump intensity. The dashed line at approximately 20 mW/cm^2 is the divider between the quasi transient and the steady state transition as shown in the figure. Each data point on the graph is the peak diffraction efficiency value corresponding to each pump intensity. Data on the left hand side of the dashed line belongs to quasi-transient efficiencies, and the right hand side is the data of the steady state diffraction efficiencies. The efficiency of a steady state diffraction lessens for a strong pump power, because the strong power saturates the gratings making them less sinusoidal in structure. Therefore, the correct ratio between write and pump power is needed to optimize the steady state efficiency. It is also a

remarkable coincidence that the diffraction efficiency as a function of pump intensity exhibits a hump behavior.
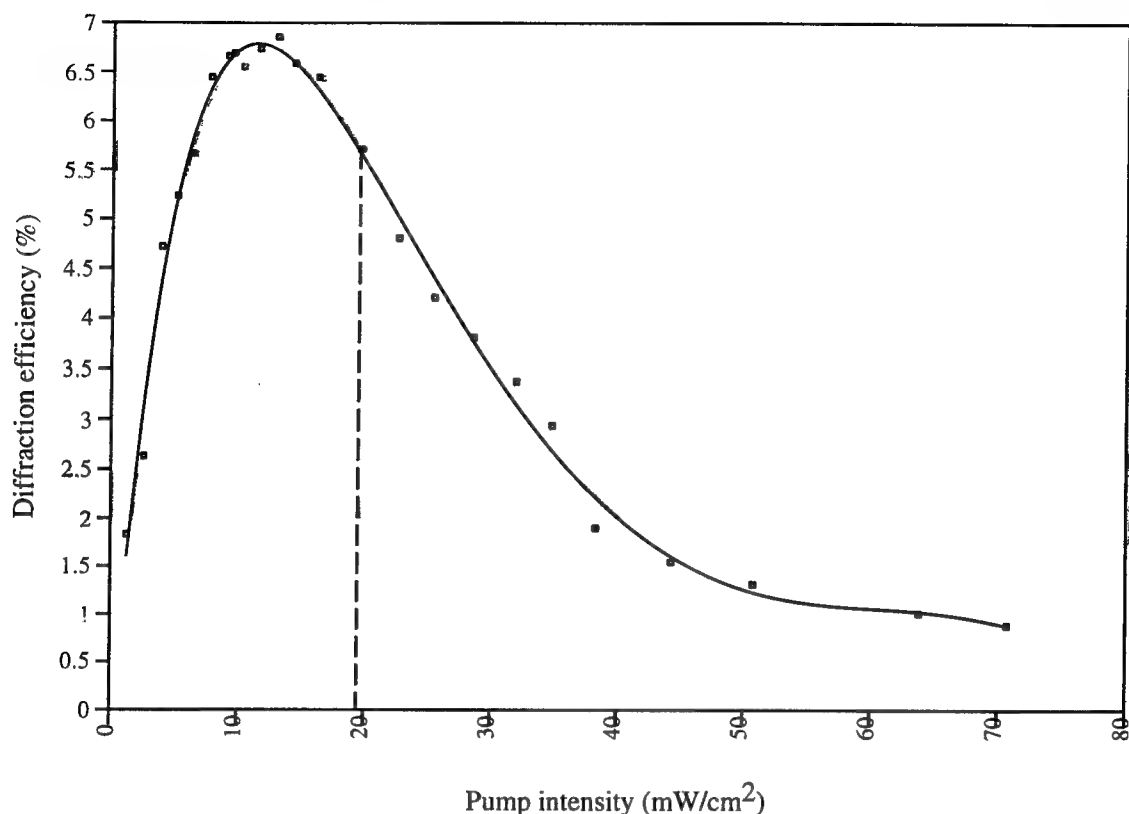
Fig. 5-13 Bragg diffraction efficiency as a function of pump intensity.

Figure 5-14 depicts rise and erase time as a function of pump intensity. The rise time for quasi-transient curves is faster than steady state curves as shown by plot a. In other words, a strong pump intensity produces a slow rise time. Plot a seems to settle down asymptotically as the intensity exceeds $50 \, mW/cm^2$. Two conclusions can be drawn from studying this plot; a strong pump power induces a slow hologram and the slowest rise time is less than 1.5 seconds for this particular sample. On the other hand, the erase time shortens with each increment of pump intensity as shown by plot b. Holographic gratings in a dynamic BR film are erasable with blue light illumination to induce a direct photo-transition from M to B-state. However, the experiment has shown that gratings can also be erased by perturbing molar concentrations in the M-state. The experiment to achieve this

task was done by blocking out one of the write beams and observing grating decay behavior for various pump intensities. The result revealed that erased time also depends on pump intensity since it is faster as the intensity increases.
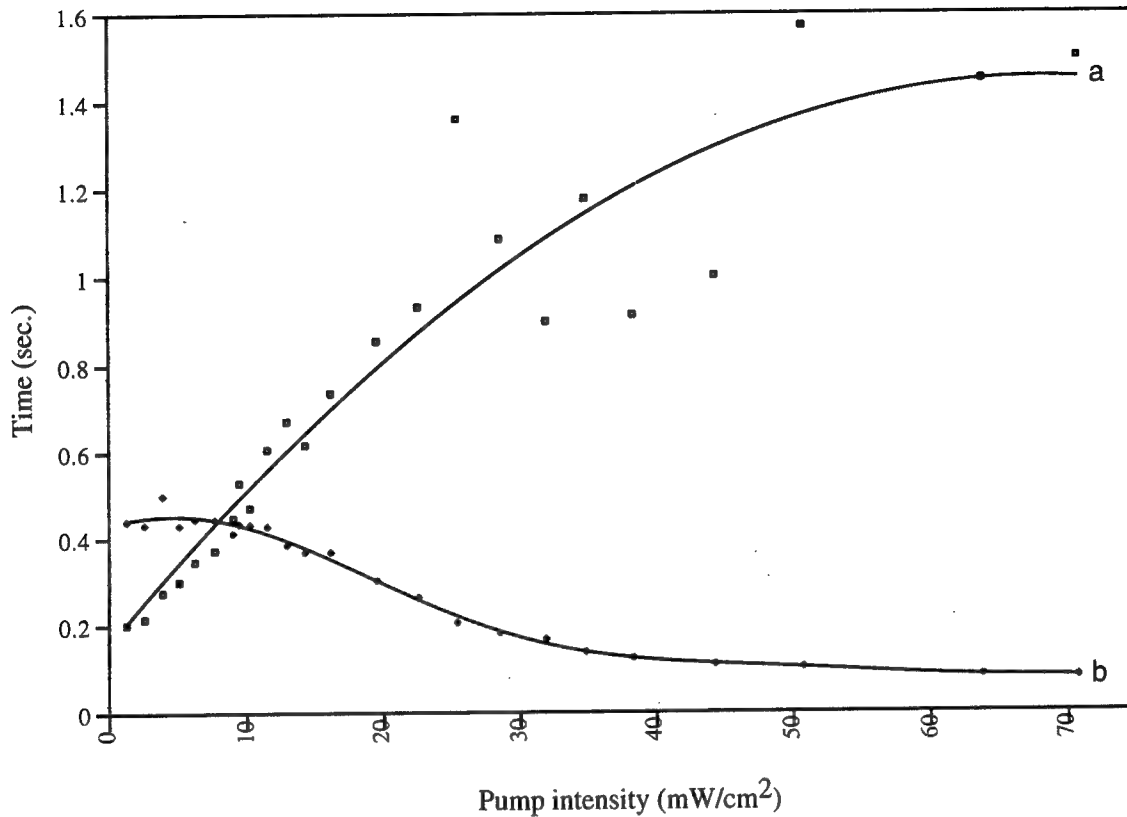
Fig. 5-14 Rise and erase time as a function of pump intensity.

Illustrations as shown in Fig. 5-15 provides a different perspective on Bragg diffraction behavior. It depicts the efficiency as a function of write intensity under a constant pump intensity. The objective of the experiment is to study this behavior by changing write intensity and keeping the pump intensity fixed at 8.15 mW/cm^2. The figure consists of steady state diffraction curves for low write intensities as shown by plots a, b and c, and a quasi transient curve for a stronger write intensity depicted by plot d. The results seem to have a reversed behavior with respect to the plots in Fig. 5-12. Thus, both figures undoubtedly concede to have efficiency higher in the quasi transient mode versus

that of a steady state condition. In addition, the strong write intensity appears to compress the hump width, which also contradicts the previous results.
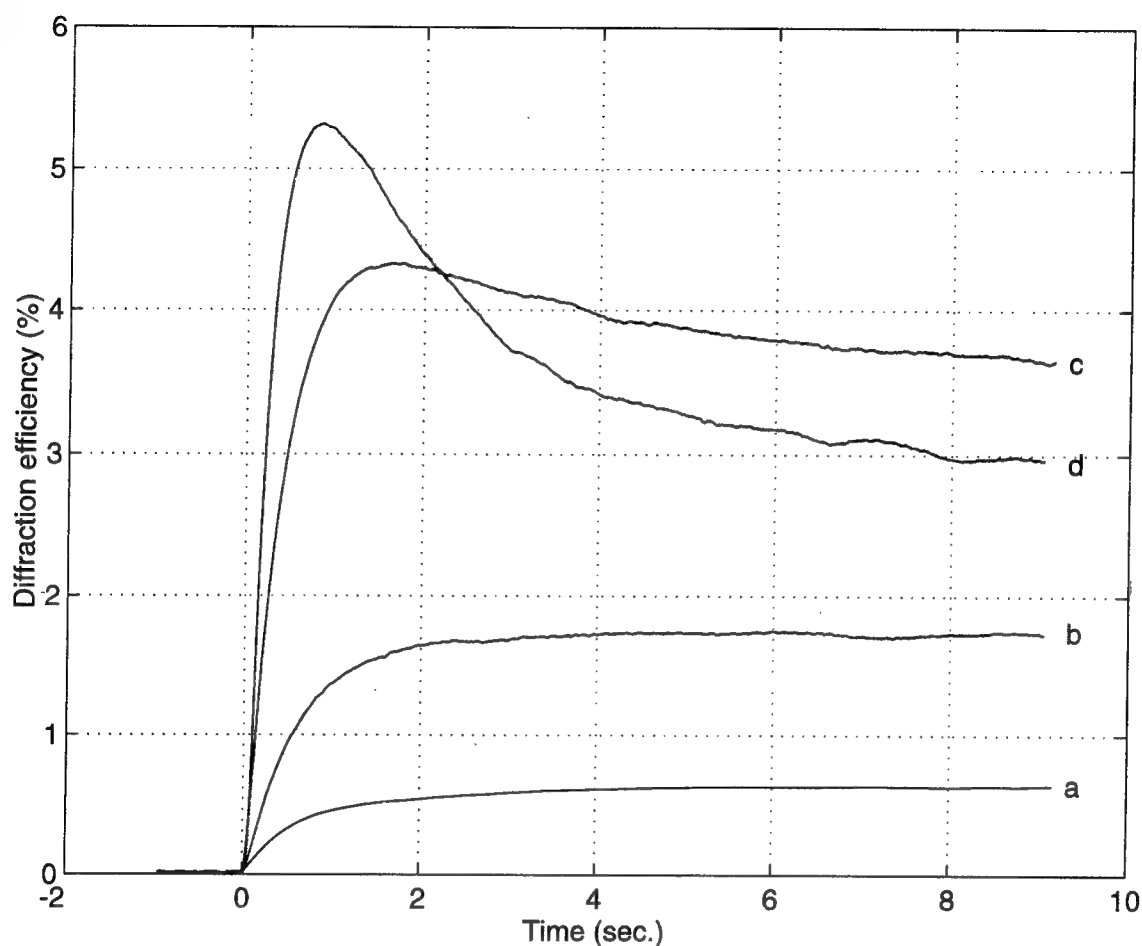
Fig. 5-15 Diffraction efficiencies for various write intensities. The intensities in mW/cm^2 are 4.91, 9.73, 22.21 and 39.32 for plots a, b, c and d, respectively.

The diffraction efficiency as a function of write intensity is shown in Fig. 5-16. The figure can be viewed as two sets of efficiency data, steady state and quasi-transient, which are partitioned by a dashed line where the intensity is approximately 22 mW/cm^2. Steady state efficiencies fall on the left hand side of the partition, with the right hand side being quasi-transient efficiencies. The efficiency increases with increase of the write intensity until it reaches a saturated value of 6 % and the intensity exceeds 40 mW/cm^2. Further increasing the intensity to 120 mW/cm^2 will not increase the peak efficiency but it

will, however, shorten the holographic rise time, as given in the next figure. The explanation may lie on refractive-index modulation coefficients which increase proportionally to the write intensity, and so too the efficiency.
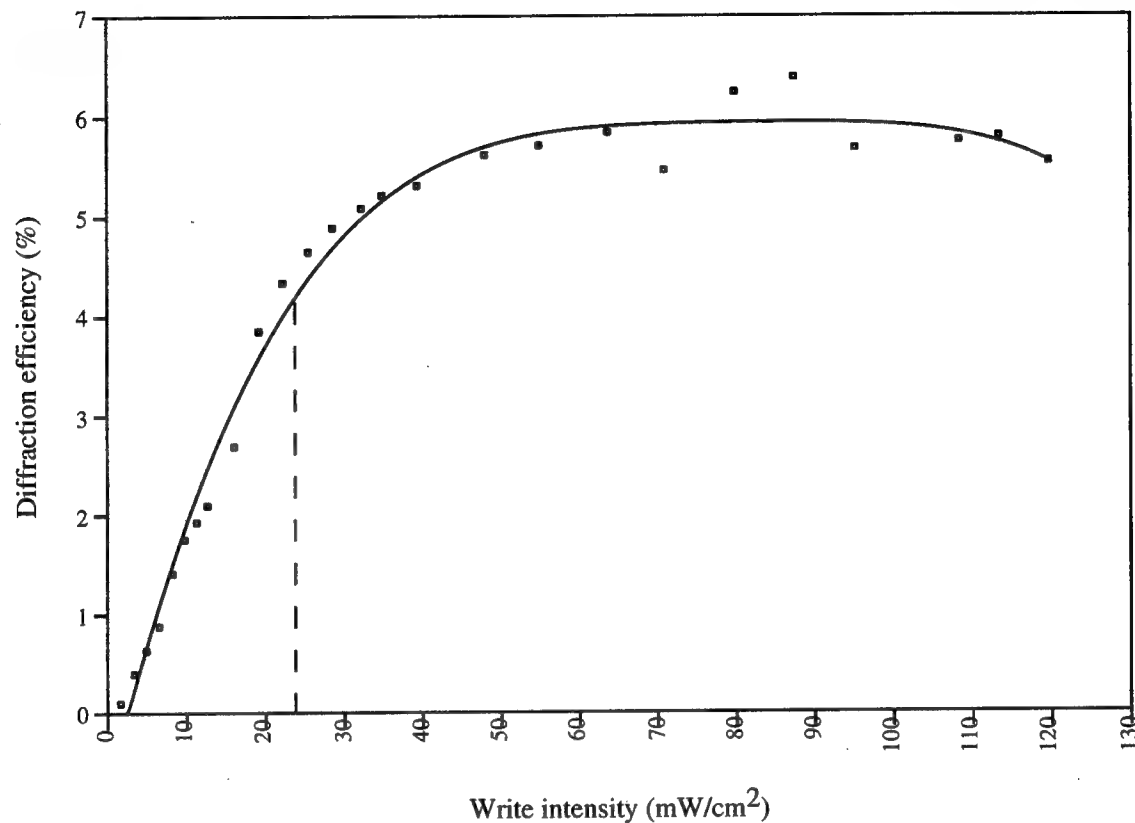
Fig. 5-16 Diffraction efficiency as a function of write intensity.

Figure 5-17 depicts the rise and erase time of M-type holograms as a function of write intensity with a fixed pump power. Rise time is longer in a steady state condition than in a quasi-transient mode. In both cases the time is shortened by a stronger write intensity. Rise time is represented by plot a, and plot b is erase time. In the experiment, gratings were erased by blocking one of the writing beams. The figure shows that increasing intensity will decrease rise and erase time. Hence, both times are intensity dependent. They begin to converge at an intensity exceeding 50 mW/cm^2, and they continue to decrease at the same rate to 100 msec. at an intensity of approximately 120 mW/cm^2. These times could be lower if the intensity is higher, especially the erase time.
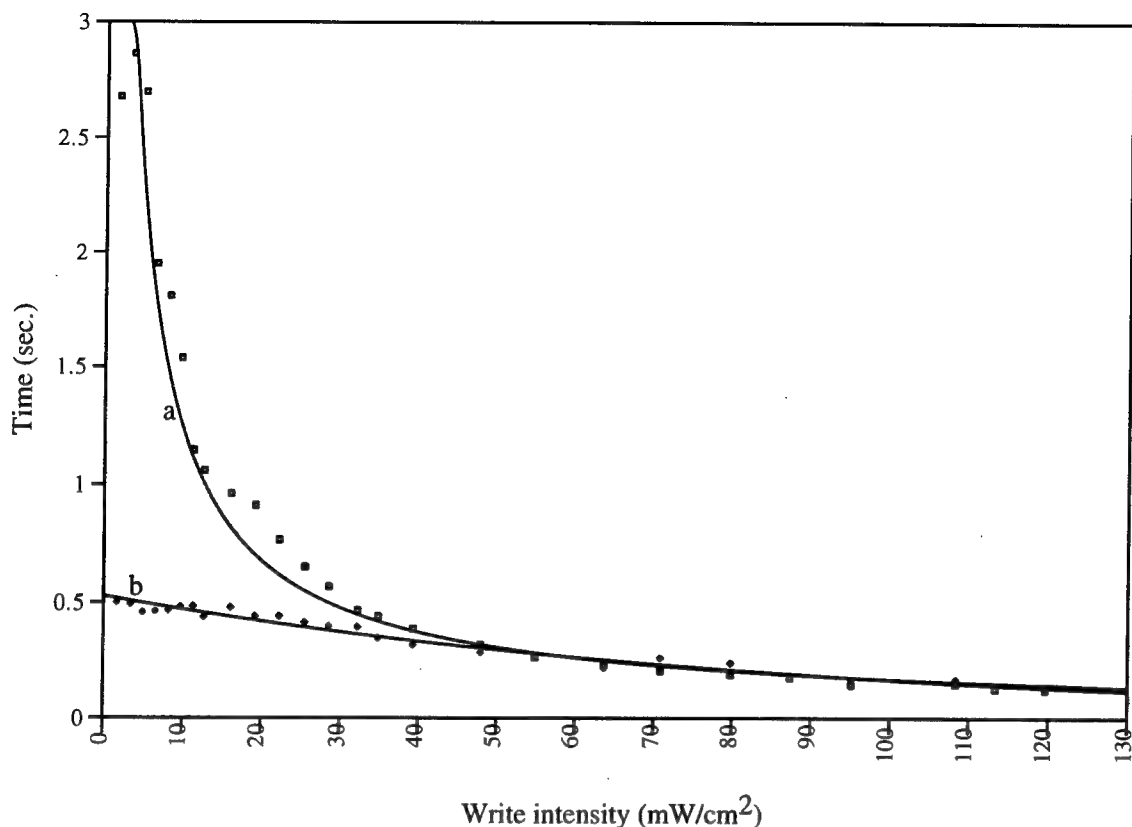
Fig. 5-17   Rise and erase time as a function of write intensity.

5.4   Transient dynamic hologram

The three laser beam technique developed in the previous section is also applicable to transient investigation by maximizing the diffraction efficiency of M-type phase holograms in a BR film. The concept is to induce a direct phototransition from the excited M-state to the B-state by the writing process with an M-state absorption laser. This method capitalizes on the bilevel BR system. Depleting the B-state creates molar concentrations in the excited M-state, and vice versa. Hence, illuminating the film with a B-state absorption laser will build up the molecules in the M-state, so that M-type holographic gratings can be written by a 413-nm wavelength laser. Therefore, the experimental setup for transient dynamic holograms is very much similar to Fig. 5-11, except SH2 is triggered open by the

130

closing of SH1. Thus, the transient diffraction profile should be a pulse-like shape with short duration as given in the next figure. The theoretical explanation discussed below will open the door to understand transient holographic diffraction gratings in a BR film.

Before the optimal exposure time, the grating absorption of the write beams enhances the refractive index modulation coefficients and deepens the modulation depth, in which the diffraction efficiency is increasing. The point of maximum efficiency marks the optimal grating amplitude and most of the sinusoidal grating structure. After that time, the writing process drives the molecules back to the ground state (B-state). Hence, the transient holographic characteristics depend on the write intensity. This process offers the advantages of better modulation and faster response times. Note that the characteristics are normally referred to diffraction efficiency and response times, and the response times include rise and decay time. Exposure time used in this section is referred to a pump time before write beams are triggered open by the termination of a pumping beam.

The purpose of this experiment is to investigate transient holographic grating characteristics with respect to exposure time, pump and write intensity. Hence, the investigation has three experimental parts of the BR4. Consequently, the strategy is to divide the investigation into three experimental parts.

The first part of the investigation was to study the characteristics as a function of exposure time for a constant pump and write power. The intensity of the write beams was kept at 31.22 mW/cm^2, while the pump intensity was fixed at 33.54 mW/cm^2 for the entirety of the experiment. By controlling a shutter's open duration of the pump beam to various exposure times, different efficiency peaks can be obtained as shown in Fig. 5-18. The figure illustrates a transient holographic diffraction pattern as a quick rise time, followed by an exponential decay. The decay provides insight as to how transient gratings are partially erased by the slow depopulation of molar concentrations in the excited M-state. Hence, the erase time is independent of the exposure time as shown in Fig. 5-19, and is approximately 2.35 seconds for this BR4 with the fixed write power level. However, the

131

peak efficiency and rise time increase with prolonged exposure time until the efficiency reaches 6.2 %. The corresponding rise and exposure time were 0.25 seconds and 1.5 seconds, respectively. Further prolonging the exposure duration will not augment the efficiency nor increase the rise time. This statement is supported by the data that runs parallel in the figure after the efficiency has reached its maximum value. The reason is the given film and pump intensity, 1.5 seconds is enough to pump all the molecules to M-state, and increasing the exposure time will not provide more molecules to the state. Results from the next experiments will reinforce on this explanation.
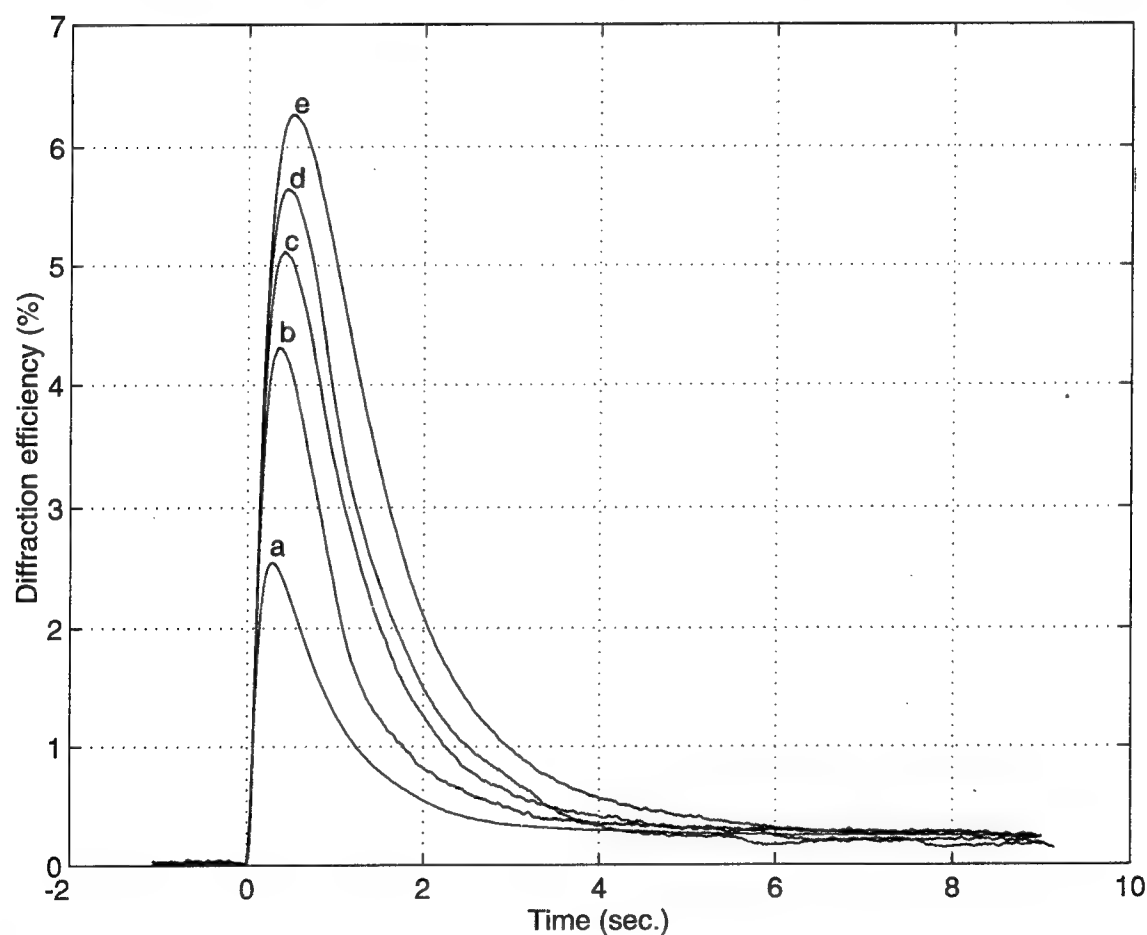
Fig. 5-18 Transient holographic diffraction efficiency profile for various exposure times. The exposure times in second are 0.08, 0.15, 0.25, 0.40 and 1.5 for plots a, b, c, d and e, respectively.
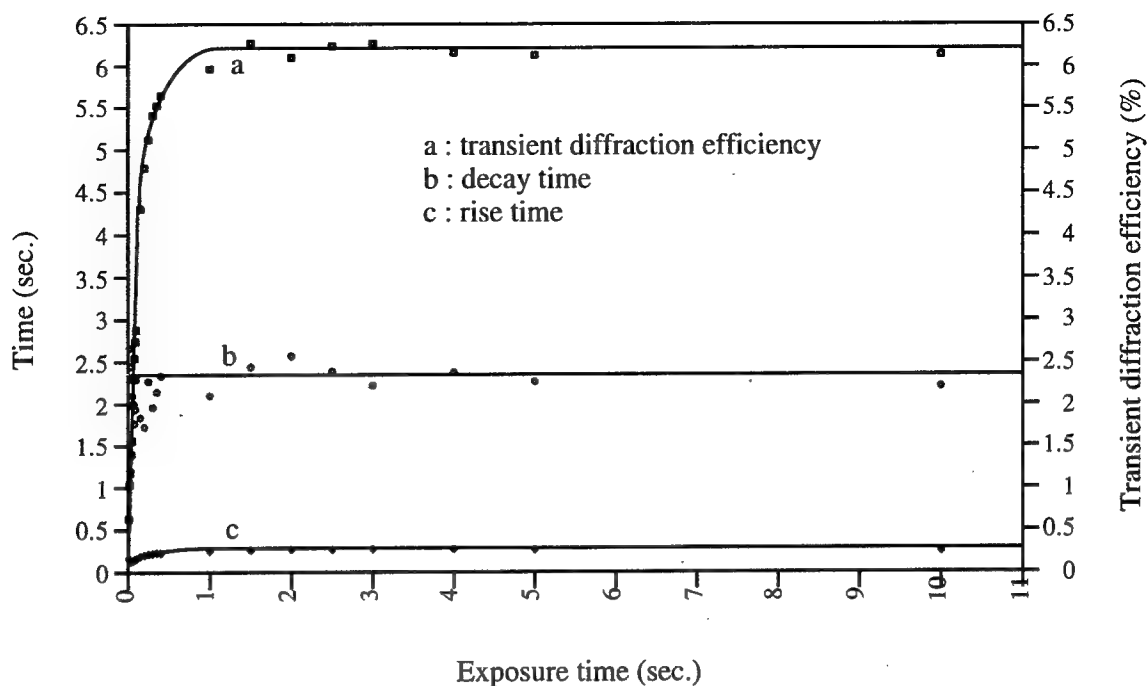
Fig. 5-19 Transient holographic grating characteristics as a function of exposure time.

This second part of the experiment is to investigate transient holographic grating characteristics as a function of write intensity, as shown in Fig. 5-20 for a pump intensity of 33.53 mW/cm² and exposure time of 2.0 seconds. The maximized transient peak diffraction efficiency increases proportionally with the write intensity when it is less than 30 mW/cm². This is because increasing the write intensity deepens modulation depth and heightens the depleting modulation amplitude in the M-state. Note that the intensity in this region is below the pump intensity. The efficiency reaches its first maximum value when the ratio of write to pump intensity approaches unity. The peak efficiency saturates at 6.6% after the ratio exceeds unity. Further increasing the write intensity up to 140 mW/cm² will not increase the peak efficiency but it will, however, shorten both response times. It stands to reason that a strong write intensity will create a large grating amplitude at an accelerated rate. The rise time approached 100 msec. as the erase time was reduced to 1 second at the maximum write intensity. The erase time is not a conventional erasure with a single blue beam. This time is a measurement of transient gratings amplitude receded by the writing

133

process. Furthermore, the 413-nm wavelength beam is one of the wavelengths at the lower end B-state absorption spectrum. Hence, writing with this wavelength may also help maintain an activity in the bilevel system by pumping the molecules back and forth between the two states. This may contribute to a gradual erasing.
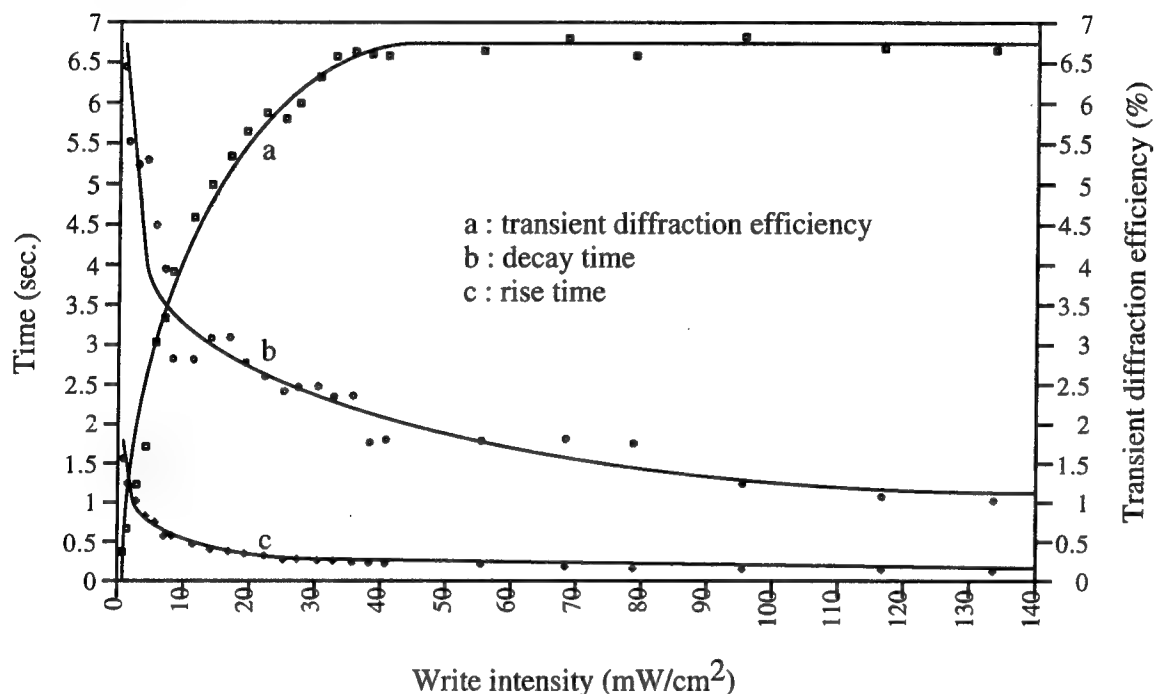
Fig. 5-20 Transient holographic grating characteristics as a function of write intensity.

Figure 5-21 depicts transient holographic grating characteristics as a function of pump intensity for a write intensity and exposure time fixed at 31.22 mW/cm^2 and 2 seconds, respectively. The peak efficiency increases with the increase of pump intensity until it reaches the write intensity value. The efficiency is maximum when the ratio of write to pump intensity is approximately unity. The peak efficiency approaches a saturated value of 6.4 % after the pump intensity exceeds 30 mW/cm^2. Pumping with stronger intensity will not achieve a higher peak efficiency nor will it change the response times. The erase time is about 2.4 seconds for this BR film under the constant write intensity. The rise time increased proportionally with the increase of efficiency. Once the ratio decreases below

unity, the rise time also saturates. In other words, the peak efficiency stabilizes around 6.4 % and the rise time slowly approaches 0.25 seconds. Note that the characteristics presented in this figure are almost a duplication of those in Fig. 5-19. A common thread between both experiments was the fixed write intensity value.
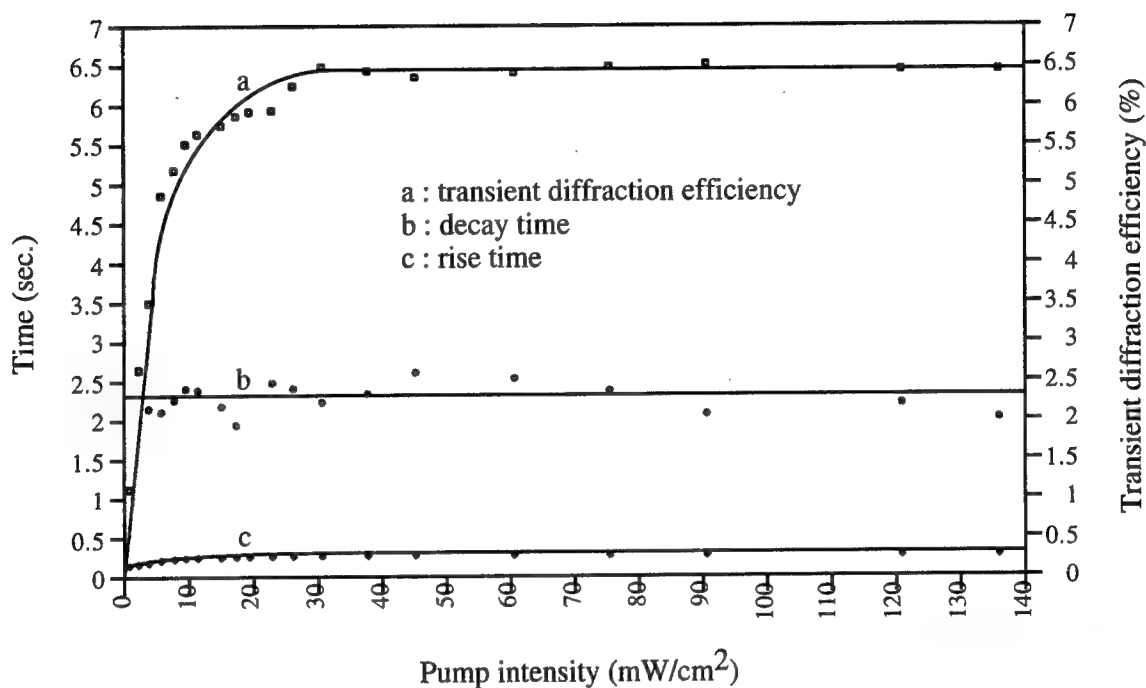
Fig. 5-21 Transient holographic grating characteristics as a function of pump intensity.

5.5 Summary

This chapter began with the experiment using a normal incidence. The purpose of this experiment is to examine BR properties in terms of response time, transmission coefficient, reflection coefficient and total absorption with respect to a 532-nm and a 680-nm light sources. Four BR films were subjected to the experiment of a normal incidence and B-type hologram. The B-state absorption laser beam was used to write holographic gratings. The reading is done with 680-nm wavelength beam which is far away from the

absorption bands. The experimental results were in favor of the hydrated BR_{D96N}. It is one of the highest diffraction efficiencies reported to date, and its response time is on the order of a millisecond. This mutated variant was further used in the experiment of the M-type and transient dynamic holograms. The commonality between these two holograms is the M-state holographic gratings. Both experiments require a pump beam to build up molar concentration in the excited M-state before gratings can be written by the M-state absorption laser beam. The M-type experiment produced a steady state diffraction profile and quasi-transient behavior. Finally, the maximized transient peak diffraction efficiency was almost 7% and the rise time was less than 200 msec.

Chapter 6

Concluding Remarks

This report has introduced a novel high capacity free-space switching fabric with dynamically reconfigurable holograms. It provides an enabling element for a proposed system that is composed of a physical link and routing control layer of the OSI hierarchical structure. Node implementation of the system was discussed in a full detail. An operational function of each component in the crossbar switching fabric was clearly described. The description explains the fundamental concept that the ith row in an Nx1 input network can be mapped to any port in the ith row of an output network. This switching device offers the virtue of an experimental simplicity. It basically consists of a light source, SLM, cylindrical lens, dynamically reconfigurable photorefractive polymer and output network. Therefore, it has the potential for a low weight and portability. A precise and accurate functional description of each processor in the routing controller was provided in a logical manner. This switching device resolves optical-to-electronic and electronic-to-optical conversion bottlenecks which is one of the major disadvantages of the currently used electronic counterpart. Furthermore, it not only reduces signal-to-noise degradation that is due to the conversion but also maintains fiber optic bandwidth capacity and induces transparent data redistribution. It also has advantages over a guided optical crossbar since it is a monostage system. Hence, the device is completely free of internal blocking. Signal degradation of its guided counterpart is compounded from stage to stage that could tremendously increase bit error rate with the increasing guided network size. Note that the hardware complexity of the proposed system is on the order of N as compared to that of a guided crossbar switch which is on the order of N^2 crosspoints, where N is the size of input-output array. In fact, a guided switch is a suitable application only for a small scale network. However, the proposed system capitalizes on the

parallelism and multidimensionality inherent in optics and can be scaled to a large interconnectivity density.

The free-space, broadband crossbar switch is based on a bio-optical material called bacteriorhodopsin (BR). The models of a reversible reaction scheme, one-photon and two-photon process for a complex photochemical reaction in a BR film were theoretically investigated. The investigations have led to the development of a three-laser beam technique that makes it possible to study holographic diffraction characteristics of M-type phase holograms and transient gratings. A B-state absorption light source pumps BR molecules from the B-state to the excited M-state, then the M-state absorption beam is used to write holographic gratings in that state, and the reading is done with a light source that is outside the absorption band to ensure a negligible erasing effect on the gratings written in the film. The experimental results have shown that response times are in the order of milliseconds, and Bragg diffraction efficiency is close to 10 % for a hydrated D96N mutant. In a switching application, attenuation of an optical signal at each node can be compensated by a fiber amplifier. Thus, one application scenario of the system is in the area of a virtual circuit packet switching scheme that is a hybrid between circuit and packet switching.

A slow switching time on the order of milliseconds limits the device's utility in circuit network applications. This time includes a routing setup time and a response time of a photorefractive polymer. The setup time improves proportionally with the increasing of a SLM speed. It would be idealistic to control routing data optically. This would completely eliminate the concern of the speed. Hence, optical controller is one approach to enhance the performance of the device. However, an entirely different switching architecture may evolve. On the other hand, the response time depends on a polymer's characteristic. It is more important than diffraction efficiency because the signal at each redistribution node can be amplified by a number of techniques. If the polymer possesses a high efficiency property then the design becomes cost-effective system. Note that the device can be used

as a packet switching network if and only if the switching time is faster than packet arrival rate. In conclusion, the speed of a SLM or optical routing controller, and the issue of improving the holographic grating characteristics of the polymer are the challenges that next need attention.

Bibliography

1. W. Stallings, Data and Computer Communications, 2nd Ed., Macmillan, New York, 1988.

2. J. C. Palais, Fiber Optic Communications, 2nd Ed., Prentice Hall, New Jersey, 1988.

3. J. Hayes, Fiber Optics Technician's Manual, Delmar, New York, 1996.

4. J. Hecht, "Low-loss fiber opens optical communications," Laser Focus World, pp. 55-57, Nov. 1994.

5. F. M. Mims III, "Alexander Graham Bell and the Photophone: The Centennial of the Invention of Light-Wave Communications, 1880-1980," Opt. News, Vol. 6, No. 1, pp. 8-16, 1980.

6. The Encyclopedia Americana International Edition, Vol. 26, pp. 399-411, Grolier, Danbury, CT., 1994.

7. J. Y. Hui, Switching and Traffic Theory for Integrated Broadband Networks, Kluwer Academic, Boston, 1990.

8. The Encyclopedia Americana International Edition, Vol. 7, pp. 490-493, Grolier, Danbury, CT., 1994.

9. S. D. Personick and W. O. Fleckenstein, "Communications Switching-From Operators to Photonics," Proc. of The IEEE, Vol. 75, No. 10, pp. 1380-1403, Oct. 1987.

10. E. Braun and S. Macdonald, Revolution in Miniature, Cambridge University Press, Cambridge, 1978.

11. K. Noda, Ed., Optical Fiber Transmission, Vol 6, Elsevier Science, Amsterdam, 1986.

12. T. H. Maiman, "Stimulated Optical Radiation in Ruby," Nature, pp. 493-494, Aug. 6, 1960.

13. G. Keiser, Optical Fiber Communications, McGraw-Hill, New York, 1983.

14. A. Javan, W. R. Bennett, Jr., and D. R. Herriott, "Population Inversion and Continuous Optical Maser Oscillation in a Gas Discharge Containing a He-Ne Mixture," Phys. Rev. Letters, Vol. 6, No. 3, pp. 106-110, Feb. 1, 1961.

15. M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill, Jr., and G. Lasher, "Stimulated Emission of Radiation from GaAs p-n Junctions," Appl. Phys. Letters, Vol. 1, No. 3, pp. 62-64, Nov. 1, 1962.

16. R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, "Coherent Light Emission from GaAs Junctions," Phys. Rev. Letters, Vol. 9, No. 9, pp. 366-368, Nov. 1, 1962.

17. T. M. Quist, R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWhorter, and H. J. Zeigler, "Semiconductor Maser of GaAs," Appl. Phys. Letters, Vol. 1, No. 4, pp. 91-92, Dec. 1, 1962.

18. Y. Suematsu and K. Iga, introduction to optical Fifer Communications, Wiley, New York, 1982.

19. K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," Proc. IEE, Vol. 113, No. 7, pp. 1151-1158, July 1966.

20. F. P. Kapron, D. B. Keck, and R. D. Maurer, "Radiation Losses in Glass Optical Waveguides," Appl. Phys. Letters, Vol. 17, No. 10, pp. 423-425, Nov. 15, 1970.

21. R. G. Hunsperger, Ed., Photonic Devices and Systems, Dekker, New York, 1994.

22. R. Pierret, Semiconductor Fundamentals, Vol. 1, Addison-Wesley, Reading, MA, 1983.

23. A. G. Bell, "Professor A. G. Bell on Selenium and the Photophone," The Electrician, pp. 214-215, Sept. 18, 1880.

24. S. Thai and D. Garafalo, "Parallel Modulation for High Speed Communications," SPIE, Vol. 2149, pp. 6-11, Jan. 15, 1994.

25. D. J. Griffiths, Introduction to Electrodynamics, 2nd Ed., Prentice Hall, Englewood Cliffs, NJ., 1989.

26. D. K. Cheng, Field and Wave Electromagnetics, 2nd Ed., Addison-Wesley, Reading, MA., 1989.

27. H. A. Haus, Waves and Fields in Optoelectronics, Prentice-Hall, Englewood Cliffs, NJ., 1984.

28. H. Kogelnik and R. V. Schmidt, "Switched directional couplers with alternating $\Delta\beta$," IEEE J. of Quantum Electronics, Vol. QE-12, No. 7, pp. 396-401, July 1976.

29. B. E. A. Saleh and M. C. Teich, Fundamentals of Photonics, Wiley, New York, 1991.

30. A. Yariv and P. Yeh, Optical Waves in Crystals, Wiley, New York, 1984.

31. L. Rebouta, J. C. Soares, M. F. Da Silva, J. A. Sanz-Garcia, E. Dieguez and F. Agullo-Lopez, "Dopants in $LiNbO_3$: lattice site location, ion implantation and epitaxial regrowth," Mat. Res. Soc. Symp. Proc., Vol. 244, pp. 311-316, Materials Research Society, Pittsburgh, Pennsylvania, 1992.

32. C. Partridge, Gigabit Networking, Addison-Wesley, Reading, Massachusetts, 1994

33. T. J. Cloonan, "Free-space optical implementation of a feed forward crossbar network," Applied Optics, Vol. 29, No. 14, pp. 2006-2012, 10 May 1990.

34. L. McCaughan and G. A. Bogert, "4x4 Ti:LiNbO$_3$ integrated-optical crossbar switch array," Appl. Phys. Lett., Vol. 47, No. 4, pp. 348-350, 15 August 1985.

35. K. E. Batcher, "Sorting networks and their applications," Proc. AFIPS, Vol. 32, pp. 307-314, Spring Joint Computer Conference, 1968.

36. H. S. Stone, "Parallel processing with the perfect shuffle," IEEE Transactions on Computers, Vol. C-20, No. 2, pp. 153-161, February 1971.

37. M. Murdocca and T. J. Cloonan, "Optical design of a digital switch," Applied Optics, Vol. 28, No. 13, pp. 2505-2517, 1 July 1989.

38. S. Thai and D. Garafalo, "High speed signal routing holographic switching system," SPIE, Vol. 2216, pp. 91-98, 4-5 April 1994.

39. P. R. Prucnal, "Optically processed self-routing, synchronization, and contention resolution for 1-D and 2-D photonic switching architectures," IEEE J. Quant. Electronics, Vol. 29, No. 2, pp. 600- 612, February 1993.

40. N. Sriskanthan, P. K. K. Loh, K. H. Lee, and Y. C. Chan, "A real-time PC-based video phone system on ISDN/LAN," IEEE Trans. Cons. Electronics, Vol. 41, No. 2, pp. 332-342, May 1995.

41. M. Ebihara, "Still video phones bring people face-to-face," AEU, No. 3, pp. 38-41, 1990.

42. R. B. Bellman, "I want a videophone-now!," Bus. Comm. Rev., pp. 38-41, Apr 1995.

43. S. Thai and J. Malowicki, "Free-space dynamically reconfigurable bacteriorhodopsin holographic crossbar," SPIE, Vol. 3230, pp. 2-11, 2-3 November 1997.

44. S. Thai, J. Malowicki and Q. W. Song, " Bacteriorhodopsin Optical Switch," SPIE, 13-17 April, 1998.

45. K. Ema, M. Kuwata-Gonokami and F. Shimizu, "All-optical sub-Tbits/s serial-to-parallel conversion using excitonic giant nonlinearity," Appl. Phys. Lett., Vol. 59, No. 22, pp. 2799-2801, 25 November 1991.

46. K. Kitayama, "Ultrafast photonic asynchronous transfer mode switch based upon parallel signal processing," Opt. Rev. Sample Issue, pp-1-6, 1994.

47. R. S. Sandige, Modern digital design, McGraw-Hill, pp. 239-246, New York, 1990.

48. W. I. Fletcher, An engineering approach to digital design, pp. 92-100, Prentice_hall, New Jersey, 1980.

49. R. J. Collier, C. B. Burckhardt and L. H. Lin, Optical Holography, pp. 6-14, 228-233, 246-250, 464-467, Academic Press, New York, 1971.

50. M. C. Hutley, Diffraction Gratings, pp. 23-26, Academic Press, New York, 1982.

51. H. J. Caulfield (Ed), Handbook of Optical Holography, pp. 29-41, Academic Press, San Diego, 1979.

52. D. M. Burland and C. Brauchle, "The use of holography to investigate complex photochemical reactions," J. Chem. Phys., Vol. 76, No. 9, pp. 4502-4512, 1 May 1982.

53. A. Vanderlugt, Optical Signal Processing, pp. 134-141, Wiley-Interscience, New York, 1992.

54. P. P. Banerjee and T. C. Poon, Principles of Applied Optics, pp. 190-213, Aksen Associates, Homewood, IL, 1991.

55. W. R. Klein, "Theoretical efficiency of Bragg devices," Proc. IEEE, Vol. 54, pp. 803-804, May 1966.

56. L. E. Hargrove, "Optical effects of ultrasonic waves producing phase and amplitude modulatiom," J. Acoust. Soc. Am., Vol. 34, No. 10, pp. 1547-1552, Oct. 1962.

57. W. R. Klein, C. B. Tipnis, and E. A. Hiedemann, "Experimental study of Fraunhofer light diffraction by ultrasonic beams of moderately high frequency at oblique incidence," J. Acoust. Soc. Am., Vol. 38, No. 2, pp. 229-233, August 1965.

58. R. Magnusson and T. K. Gaylord, "Diffraction regioms of transmission gratings," J. Opt. Soc. Am., Vol. 68, No. 6, pp. 809-813, june 1978.

59. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. Vol. 48, No. 9, pp. 2909-2947, Nov. 1969.

60. R. R. Birge, "Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin," Biochim. Biophys. Acta, Vol. 1016, pp. 293-327, 1990.

61. R. R. Birge, "Protein-based three-dimensional memory," American Scientist, Vol. 82, pp. 348-355, July-August 1994.

62. W. Stoeckenius, "The purple membrane of salt-loving bacteria," Scientific American, Vol. 234, No. 6, pp. 38-46, June 1976.

63. D. Oesterhelt and W. Stoeckenius, "Isolation of the cell membrane of Halobacterium halobium and its fractionation into red and purple membrane," Methods Enzymol, Vol. 31, pp. 667-678, 1974.

64. J. D. Downie and D. T. Smithey, "Measurements of holographic properties of bacteriorhodopsin films," Applied Optics, Vol. 35, No. 29, pp. 5780- 5789, 10 October 1996.

65. T. Kouyama, K. Kinosita and A. Ikegami, "Structure and function of bacteriorhodopsin," Adv. Biophys., vol. 24, pp. 123-175, 1988.

66. S. O. Smith, A. B. Myers, J. A. Pardoen, C. Winkel, P. P. J. Mulder, J. Lugtenburg and R. Mathies, "Determination of retinal Schiff base configuration in bacteriorhodopsin," Proc. Natl. Acad. Sci. USA, vol. 81, pp. 2055-2059, April 1984.

67. D. Oesterhelt, C. Brauchle and N. Hampp, "Bacteriorhodopsin : a biological material for information processing," Q. Rev. Biophysics, Vol. 24, No. 4, pp. 425-478, 1991.

68. A. Popp, M. Wolperdinger, N. Hampp, C. Brauchle and D. Oesterhelt, "Photochemical conversion of the O-intermediate to 9-cis-retinal containing products in bacteriorhodopsin films," Biophys. Journal, Vol. 65, pp. 1449-1459, Oct. 1993.

69. N. Hampp, C. Brauchle and D. Oesterhelt, "Bacteriorhodopsin wildtype and variant aspartate-96 $\rightarrow$ asparagine as reversible holographic media," Biophys. Journal, Vol. 58, pp. 83-93, July 1990.

70. D. Zeisel and N. Hampp, "Spectral relationship of light-induced refractive index and absorption changes in bacteriorhodopsin films containing wildtype BR_{WT} and the variant BR_{D96N}," J. Phys. Chem., Vol. 96, No. 19, pp. 7788-7792, 1992.

71. N. Hampp, C. Brauchle and D. Oesterhelt, "Bacteriorhodopsin as a reversible holographic medium in optical processing," Ann. Intern. Conf. IEEE Eng. in Med. and Bio. Soc., Vol. 12, No. 4, pp. 1719-1720, 1990.

72. G. Varo and J. K. Lanyi, "Kinetic and spectroscopic evidence for an irreversible step between deprotonation and reprotonation of the Schiff base in the bacteriorhodopsin photocycle," Biochemistry, Vol. 30, No. 20, pp. 5008-5015, 1991.

73. G. Varo and J. K. Lanyi, "Distortions in the photocycle of bacteriorhodopsin at moderate dehydration," Biophys. J., Vol. 59, pp. 313-322, Feb. 1991.

74. R. R. Birge, T. M. Cooper, A. F. Lawrence, M. B. Masthay, C. Vasilakis, C. F. Zhang and R. Zidovetzki, "A spectroscopic, photocalorimetric, and theoretical investigation of the quantum efficiency of the primary event in bacteriorhodopsin," J. Am. Chem. Soc., Vol. 111, No. 11, pp..4063-4074, 1989.

75. D. Haronian and A. Lewis, "Elements of a unique bacteriorhodopsin neural network architecture," Appl. Opt., Vol. 30, No. 5, pp. 597-608, 10 Feb. 1991.

76. R. R. Birge and C. F. Zhang, "Two-photon double resonance spectroscopy of bacteriorhodopsin. Assignment of the electronic and dipolar properties of the low-lying $^1A_g^*$-_like and $^1B_u^*$+_like π, π^* states," J. Chem. Phys., Vol. 92, No. 12, pp.7178-7195, 15 June 1990.

77. R. R. Birge, "Photophysics and molecular electronic applications of the rhodopsins," Annu. Rev. Phys. Chem., Vol. 41, pp. 683-733, 1990.

78. R. Thoma, N. Hampp, C. Brauchle and D. Oesterhelt, "Bacteriorhodopsin films as spatial light modulators for nonlinear-optical filtering," Opt. Lett., Vol. 16, No. 9, pp. 651-653, 1 May 1991.

79. R. R. Birge, D. S. K. Govender, R. B. Gross, A. F. Lawrence, J. A. Stuart, J. R. Tallent, E. Tan and B. W. Vought, "Bioelectronics, three-dimensional memories and hybrid computers," International Electron Devices Meeting, San Francisco, CA, 11-14 Dec. 1994.

80. R. R. Birge, Protein-based branched-photocycle three-dimensional optical memories, Final Technical Report, RL-TR-96-274, 1997.

81. M. Weissbluth, Photon-Atom Interactions, pp. 223-229 and 370-376, Academic Press, Boston, MA, 1989.

82. A. Miller and D. Oesterhelt, "Kinetic optimization of bacteriorhodopsin by aspartic acid 96 as an internal proton donor," Biochimica et Biophysica Acta, Vol. 1020, pp. 57-64, 1990.

83. C. K. Mathews and K. E. van Holde, Biochemistry, The Benjamin / Cummings, pp. 42, Redwood City, CA, 1990.

84. Q. W. Song, S. Thai and J. Malowicki, "The effect photochromic nonlinearity of dried blue membrane bacteriorhodopsin films," OSA, MAAA4, pp. 97, Rochester, NY, October 20-24, 1996.

85. C. H. Chang, J. G. Chen, R. Govindjee and T. Ebrey, "Cation binding by bacteriorhodopsin," Proc. Natl. Acad. Sci., Vol. 82, pp. 396-400, Jan. 1985.

86. H. Otto, T. Marti, M. Holz, T. Mogi, L. J. Stern, F. Engel, H. G. Khorana and M. P. Heyn, "Substitution of amino acids Asp-85, Asp-212, and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK of the Schiff base," Proc. Natl. Acad. Sci., Vol. 87, pp. 1018-1022, February 1990.

87. Eds: C. H. Bamford and C. F. H. Tipper, The theory of kinetics, Vol. 2, pp. 1-13, Elsevier, Amsterdam, 1969.

88. Q. W. Song, C. Zhang, R. Blumer, R. B. Gross, Z. Chen, and R. R. Birge, "Chemically enhanced bacteriorhodopsin thin-film spatial light modulator," Opt. Lett., Vol. 18, No. 16, pp. 1373-1375, August 15, 1993.

89. R. W. Boyd, Nonlinear Optics, Academic Press, pp. 14-15, 90-93, Boston, 1992.

90. O. Werner, B. Fisher, A. Lewis, and I. Nebenzahl, "Saturable absorption, wave mixing, and phase conjugation with bacteriorhodopsin," Opt. Lett., Vol. 15, No. 20, pp. 1117-1119, Oct. 15, 1990.

91. Q. W. Song, X.M. Wang, R. R. Birge, J. D. Downie, D. Timucin, C. Gary, J. Malowicki, and S. Thai, "Gaussian-beam propagation in a bacteriorhodopsin film," OIST, 26-30 Aug. 1997.

92. C. Zhang, Q. W. Song, C. Y. Ku, R. B. Gross and R. Birge, "Determination of the refractive index of a bacteriorhodopsin film," Opt. Lett., Vol. 19, No. 18, pp. 1409-1411, Sept. 15, 1994.

93 Eds: P. Gunter and J. P. Huignard, Photorefractive materials and their applications I fundamental phenomena, Vol. 61, pp. 47-52, Springer-Verlag, Berlin, 1988.